

# Psychological Bulletin

## CONTENTS

Adaptation of Insects to Pest Experience and Insecticidal Action	CARL B. ZUCKERMAN AND IRVIN ROCK	269
Designing Experiments: Validity of Experiments	DONALD T. CAMPBELL	297
The Nature of the Subject in Psychological Research	HAROLD B. MURDOCH, JR.	313
Some Effects of Alcohol on Visual Afferents	MORIJI SAGARA AND TADASU OYAMA	327
Some Problems in the Use of Correlation Coefficients in Some Generalizations	MALCOLM D. ALNOULT	339
One-Tailed Tests and One-Tailed Alternatives	HERBERT D. KNIGHT	351
On the Use of Nonparametric Techniques	NATHANIEL KLEITMAN	354
On the Use of Nonparametric Techniques: Kleitman's Note	ROBERT J. ELLINGTON	360
On the Use of Nonparametric Techniques: The Wilcoxon Test of Analysis of Variance	QUINN McNEMAR	361

Dimensions of Individual Differences by the Factor-Analytic Psychological Association

JULY, 1957

**Wayne Dennis Bellamy**  
**Brooklyn College**

**Consulting Editors**

**LAWRENCE P. CARTER**  
24-1122 Carpenter Hall  
State University of New York  
**EDWARD CARDEN**  
Brooklyn College  
**VICTOR C. KAIMAN**  
University of Colorado

**JOHN C. DILLON**  
University of Michigan  
**ROBERT D. LEHRER**  
University of California  
**JOHN W. MCGINNIS**  
University of Michigan  
**JOHN R. STUSSMAN**  
University of Michigan

**ARTHUR S. ROFFMAN**, Managing Editor

**HILARY ODE**, Assistant Managing Editor

**Editorial Staff: FRANCIS H. CLARK, BARBARA COOPER, ROBERT F. GOLDBECK,**

*The Psychological Bulletin* contains evaluations, reviews, and critical discussions on research methodology in psychology. This journal also publishes a limited number of original research or original theoretical articles.

Manuscripts should be sent to Wayne Dennis, Dennis Hall, Brooklyn College, Brooklyn 16, New York.

**Preparation of articles for publication.** Authors are invited to consult the general directions given in the "Publication Manual of the American Psychological Association" (*Psychological Bulletin*, 1952, 49 [No. 4], pp. 1-19). A copy of this publication should be given to the section on the preparation of manuscripts (pp. 39-40), since this is a particular source of difficulty in manuscript preparation. All copy must be double spaced, including references. Figures should be submitted in duplicate. Original figures are preferred. Photocopies of figures may be photographic or pencil-drawn devices. It is suggested that authors send a copy of the manuscript to guard against loss in transmission. Fifty free offprints are given to contributors; additional offprints and offprints of early publication articles receive no gratis service. Requests for additional offprints— including subscriptions, orders of reprints, and changes of address—should be addressed to the American Psychological Association, 1200 17th Street N.W., Washington 6, D.C. Address changes should be sent to the Society Office by the 10th of the month to take effect the following month. Unsolicited manuscripts mailing from address changes will not be returned. It is the responsibility of the author to inform the post office that they will guarantee second-class postage. Contributors of accepted manuscripts receive payment for their work. Reprint requests should be addressed to the author.

**Annual subscription:** \$8.00 (Foreign \$8.50). Single copies \$1.00.

PUBLISHED BIMONTLY

**THE AMERICAN PSYCHOLOGICAL**

**ASSOCIATION, INC.**  
Washington, D.C.  
and 1200 Sixteenth Street N.W., Washington, D.C.

Second class postage paid at the post office at Washington, D.C., and at additional mailing offices. Postage paid for in Canada by Postmaster General, Ottawa, Ontario, Canada.

Copyright, 1953, by The American Psychological Association.

# Psychological Bulletin

A REAPPRAISAL OF THE ROLES OF PAST EXPERIENCE AND INNATE ORGANIZING PROCESSES IN VISUAL PERCEPTION<sup>1</sup>

CARL B. ZUCKERMAN

*Brooklyn College*

AND IRVIN ROCK

*Graduate Faculty of Political and Social Science, New School for Social Research*

Twenty years ago, in his *Principles of Gestalt Psychology*, Kurt Koffka posed the problem of visual perception in the succinct question, "Why do things look as they do?" He took issue with the usual answer that things look as they do because of our past experience with them, arguing that an empiristic theory not only failed to account for many of the facts of visual perception, but, in addition, entailed a number of logical difficulties.

Few of Koffka's arguments have been satisfactorily met; nevertheless, two decades later, a survey would show that his analysis has had little or no impact on the psychological literature dealing with this problem. For example, a widely used textbook states: "With the few possible exceptions provided by primitive organizations, all perceiving is dependent upon past experience—the so-called habit factor" (45, p. 410). Moreover, the empiristic theory has re-emerged in the currently popular assumption that perceptions are governed by motivational and affective forces; in this remodeling, however,

critical questions involved in an experiential approach to perception continue to be overlooked.

The present paper attempts to reconsider the logic of the central problem and to examine the evidence bearing, in particular, on the question of whether form perception is learned. We shall restrict our discussion to the controversy between theories which emphasize the role of learning (empiristic theory) and the theory which stresses the role of innate organizing processes (which we shall briefly refer to as the organization theory). The concept of organization is, of course, basic to Gestalt psychology. With respect to its relevance for the field of perception, however, it can be evaluated on its own merits quite apart from the validity of other aspects of Gestalt theory, particularly the physiological theories advanced by Gestalt psychologists.

## ANALYSIS OF THE TWO THEORETICAL APPROACHES

It is difficult to find a clear, unambiguous statement of an empiristic position; moreover, many writers assume the validity of empiristic hypotheses but do not offer an analysis of basic questions. The formulation of the problem by Ames and his co-workers (the transactional approach) may serve, however, to il-

<sup>1</sup> We take this opportunity to state our indebtedness to Dr. Hans Wallach who has so greatly influenced our approach to the problems discussed in this paper. We also wish to express our gratitude and appreciation to Dr. Evelyn Raskin for her invaluable editorial assistance.

lustrate a modern empiristic theory which has focused on some of the essential issues in the problem of perception.

The transactionalists (28) argue that the percept cannot be derived from the retinal image alone, since an infinity of external objects can give rise to the same pattern of stimulation on the retina. For example, a small object nearby and a large object at a greater distance can result in the same sized retinal image; similarly, a circular retinal image may be produced by a circle in the frontal parallel plane or by an ellipse tilted from this plane. Or, once again, a retinal image of a specific intensity may be produced by either a black object in bright illumination or a white object in dim illumination. How then, in view of this equivalence of outer configurations in producing identical retinal images, can the organism "know" which object to see?<sup>2</sup> The answer given to this question is that the explanation is to be sought in the realm of past events; the retinal stimulus pattern must be interpreted in the light of knowledge from the past.

The question of what the organism sees originally—before it is able to interpret the retinal pattern—is not raised. It seems clear, however, that an empiristic position of this kind

<sup>2</sup> It is true, of course, that a given retinal form may be produced by an "infinity" of external configurations. This statement must, however, be qualified. The same elliptical retinal image can result from an elliptical object in the frontal parallel plane or from a variety of circles at different tilts from this plane, etc., and, in this sense, the retinal image is ambiguous. But, under no circumstances, could a retinal ellipse be produced by a rectangular or triangular object. There is a limitation, then, to the ambiguity of the retinal image and, accordingly, there is no need for invoking assumptions to explain why we see a rounded object and not a triangle.

must hold that the initial perceptual experience would be ambiguous—a given image could result in the perception of a small or large object, a black or white object, etc. By means of "purposive action" with respect to the object, we build up "assumptions" which then determine the nature of the present perceptual experience.

Empiricists in the past formulated the problem in a similar way, but instead of speaking about the ambiguity of the stimulus they emphasized the fact that the stimulus is frequently such that it should lead to a percept different from the one which actually occurs. They pointed out, for example, that the shrinking image-size of an object as it moves away from the eye should result in the perception of diminishing size. Size constancy could not, therefore, be explained by the retinal image alone; the latter had to be supplemented or modified by the contributions of previous learning. Moreover, the sense of touch, rather than purposive action, was thought to provide the basis for the learning needed to attain the correct percept (especially, in the case of form perception).

Köhler (32) has pointed out that underlying the empiristic concept is the implicit assumption of the existence of a one-to-one correspondence between *local* retinal stimulation and the resulting sensory experience. Any change in the local stimulus, therefore, should result in a corresponding change in the percept. The fact that such a change does not always occur (e.g., perceptual constancies) had to be explained.

The organization theory differs from empiristic theory in its conception of the physiological correlate of the percept. Empiristic theorists have assumed that the percept should

be correlated with the process initiated by the local retinal stimulus. Organization theorists have related the percept to a more comprehensive set of central processes initiated by certain relationships in the stimulus pattern. When the stimulus is defined in relational terms, it is no longer always necessary to consider the retinal image either as inadequate or as ambiguous for the determination of the percept and to invoke past experience as a way out.

The difference between the two theories in this respect can be most clearly illustrated by reference to the problem of achromatic color perception. The empiricist would say that since different intensities of reflected light may give rise to the same percept (an object in different illuminations appears the same color—i.e., brightness constancy) or since the same intensity may give rise to different percepts (a piece of coal in bright illumination and a white paper in shadow which reflect equal amounts of light to the eye), the proximal stimulus is consequently either ambiguous or inadequate.

According to the organization theory, however, what is seen in a particular region of the visual field depends not only on the properties of the retinal image corresponding to this region (the local stimulus) but also on stimulation from adjacent or surrounding areas. Wallach (72) has clearly shown that the stimulus for achromatic surface color is not the absolute intensity of light from region A alone but is the ratio of light intensities from regions A and B. With-

out changing the intensity from A, the perceived color in A can be made to vary from black to white by changing the intensity from B. The specific neutral color seen will depend on the ratio of the two light intensities.<sup>3</sup>

The assumption of one-to-one correspondence between the local stimulus and perception is therefore invalid. When the stimulus is considered as a relational pattern, it is not ambiguous as the determinant of perceived neutral colors. The coal in bright illumination and the paper in shadow do not give rise to the same *pattern* of retinal excitation. The ratio between the intensity of the object and that of its surround would be different in each case and therefore the perceived colors would differ. Conversely, we can take an object of a particular albedo, place it on a background of some given color and vary the illumination. In spite of the changing amount of light reflected from the object, it will be seen as the same neutral color (brightness constancy) because the ratio of light intensities from the object and its background remains the same. There is, then, no necessity for assuming that the organism has to learn to see the object as black in one case or as white in the other (or as the same color in constancy situations).<sup>4</sup>

<sup>3</sup> Reflection from two surfaces represents the simplest stimulus for neutral color; in everyday life, of course, the stimulus conditions are more complex.

<sup>4</sup> As a matter of fact, careful consideration of Wallach's findings indicates that a learning theory for perceived achromatic color (or for brightness constancy) is impossible. For learning to occur, the organism would have to take into account the illumination in which a particular gray surface is given and to correct for changing illumination. There is, however, no way in which illumination can be registered independently from the surface color; both are given by the same stimulus variable



FIG. 1

The same approach is applicable to the problem of size perception. A particular retinal image may correspond either to a large object far away or a smaller object nearby. The stimulus situation is ambiguous only when distance cues are eliminated or are inaccurate. It is, therefore, entirely possible that the underlying correlate for perceived size is the interaction of the area excited in the visual cortex corresponding to the size of the retinal image and the physiological correlate of phenomenal distance (whether the distance cues themselves are learned or not).<sup>4</sup> Such an interaction process may be an outcome of learning but until this can be proven, the alternative of innate organization cannot be ruled out.

In certain cases, however, even when the percept is considered to be based upon stimulus relationships, ambiguity as to what will be perceived still persists. The following example illustrates the point. Suppose we have in a darkroom situation a luminous point *A*, surrounded by a luminous rectangle *B*. Duncker (14), in his investigation of the stimulus conditions for phenomenal movement, found that, if the rectangle *B* is slowly moved to the right, point *A* is seen to move to the left, while the rectangle is perceived as stationary. Unless the stimulus situation is defined relationally, one would have to predict that *B* would be seen to move

---

—the amount of light reflected from the object.

<sup>4</sup> Size constancy (defined functionally as a process of interaction of retinal size and perceived distance) should be distinguished from the problem of distance perception per se. Evidence that distance cues are entirely or partly learned would not prove that this interaction process is learned. Conversely, if distance cues are innate, it does not follow that size constancy is innate.

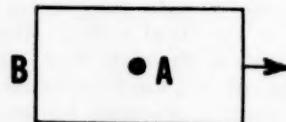


FIG. 2

since only the image of *B* is displaced on the retina. But even if defined in this way, the stimulus condition is still ambiguous, since in physical terms the situation can be correctly described either as *A* being displaced with reference to *B* or *B* in relation to *A*. Thus there are two possibilities; phenomenally, however, one is realized—only point *A* is seen to move.

How is this "preference" to be explained? Early empiristic theory would have maintained that only *B* should appear to move. Transactionalists might argue that initially either or both possibilities could be experienced and we learn to see only the point move, since in real life it is the smaller, surrounded object that usually moves. Carr has offered a similar explanation (8).

Duncker believed that seeing point *A* move is a consequence of the operation of a *selective principle* which may be defined as follows: When an object *A* is surrounded by a second object *B*, then no matter which one is actually moving, only the surrounded one will be seen to move, the outer object taking on the character of a frame of reference which tends to be perceived as stationary. So strong is this principle that, even if the surrounded object is the observer himself, he will feel himself to be in movement although objectively it is the surrounding object or scene which is moving (induced motion of the self). According to this viewpoint, the law of surroundedness represents an outcome of innate organizing factors in the

brain and not a product of learning.

It may be useful at this point to summarize the principal features of the organization theory:

1. The percept is considered to be based on stimulus relationships. Some examples of phenomena explicable in terms of relational stimulus conditions in addition to achromatic surface color (ratio of light intensities) and movement (relative displacement) are: phenomenal velocity (rate of figural change [6, 73]); geometrical illusions; chromatic color contrast; and those listed below as illustrating the operation of selective principles.

2. In many cases, it is necessary to assume, in addition to the relational properties of the stimulus, the operation of selective principles according to which sensory data are organized. Thus, one perceptual experience arises rather than another, although, on the basis of stimulus conditions, both are equally possible.\* Some examples where such principles are assumed to be operating are: laws of grouping (80); apparent movement; figure-ground organization (58); laws of surroundedness and separation of systems in movement (14); sound localization by head movements (74); depth based on retinal disparity (82);

\* Some writers, while opposing the empiricist view, do not see any need for a concept of organization. Thus, Gibson (19) and also Pratt (49) argue that it is sufficient to correlate the stimulus conditions with the resultant percept in accordance with traditional psychophysical method. Gibson has also pointed up the necessity for correlating the percept with more complex aspects of the stimulus. He does not see the need, however, to correlate the percept with the central processes initiated by the stimulus relationships, as does the organization theory. The above examples of selective principles show very clearly that the proximal stimulus itself, however defined, does not contain all that is needed for an explanation of the percept.

kinetic depth effect (77); phenomenal identity (41, 69).<sup>7,8</sup>

3. The further assumption is made that the percept is *innately* determined by such stimulus relationships and selective principles of organization. Lashley has described this position in the following statement:

"The nervous system is not a neutral medium on which learning imposes any form of organization whatever. On the contrary, it has definite predilections for certain forms of organization and imposes these upon the sensory impulses which reach it" (35, p. 35).

It is possible for an empiricist to agree with organization theory that the stimulus is a relational affair; in addition, he might even agree to the assumption of selective principles. He would then have to argue that these principles are based upon past experience. These views, however, would represent a radical change from traditional empiricist thinking; there are indications that empiricist theory may now be moving in this direction (48, 71).

The contrast between the two theoretical approaches can be further illustrated by the problem of form perception.

We assume that it is now generally agreed that the relative position of points in the visual field does not have to be learned but is given by the relative position of corresponding points of excitation in area 17 (although at

<sup>7</sup> It is still too early to tell whether phenomenal causality as investigated by Michotte (43), should be included in the above list.

<sup>8</sup> It does not seem to the authors that the concept of Pragnanz is either clear or helpful in dealing with perceptual phenomena; moreover, there is very little unambiguous evidence to support it. On the other hand, selective principles as described by Wallach do seem to imply some tendency toward preserving constancy in perceptual experience.

one time this was a debated issue). The orderly projection of the retinal points to the visual cortex in such a way as to preserve the same position of points relative to one another (topologically) seems a sufficient condition for the explanation of the perception of visual direction. Walls (79) refers to these anatomical facts as an additional argument against an empiristic theory of visual direction. Furthermore, there is evidence against a learning theory for radial direction —i.e., the direction of a point from the observer (see p. 282). A precondition for radial direction must surely involve the perception of correct position of points relative to one another. Consequently, the major unresolved issue in the area of form perception is whether *organization* of the field is a result of learning.

As experienced, the visual field is not a patchwork of various colors and brightnesses but consists of circumscribed units, certain areas belonging together and forming shaped regions which are segregated from other regions. Wertheimer (80) emphasized that segregation in the visual field was not a fact to be taken for granted but one which presented a crucial problem in the investigation of perceptual processes. One explanation of this problem has been given in terms of the retinal image. But the explanation that one sees a book because the image of a book stimulates the retina is insufficient. Sometimes one may not see a segregated unit when its image is objectively present (e.g., camouflage) and at other times, one may see a unit where objectively there is none on the retina (e.g., a star constellation). An even more fundamental objection is the fact that, although the retinal image may accurately represent the external situation, in that a homogeneously col-

ored form would give rise to an image having the correct shape and representing the color appropriately,<sup>9</sup> there is no reason why the percept should be correctly organized (i.e., in agreement with the shape and segregation of that form in the external world). The mosaic of stimuli on the retina could be organized in various ways. It is logically possible, for example, to see part of the form together with part of the surrounding area as one unit; the shape of this unit would be determined by the parts which are united.

The tendency to attribute certain aspects of perceptual experience to the retinal stimuli (the view that organized shape is given by the image) has been called by Köhler (32) the "experience error." Many empiristic writers, including Ames, do not explicitly deal with the problem of form perception, apparently not realizing that it is a problem. Similarly, S-R theorists generally speak of a form as a stimulus which is given and simply assume that no explanation is necessary. Since, however, the organized percept is not directly given by the retinal image, a theory is needed to explain the perception of forms.

The "correct" organization can be explained in two ways. Empiristic theories assume that the sensory data are structured only as a result of learning. According to this view, a young child or animal would initially not experience a visual field with segregated objects. Instead, it would see a mosaic of different brightnesses and colors (or perhaps an *incorrectly* segregated field). Murphy (46), for example, states that the infant has to learn to sort out his impressions and to learn that certain stimuli go with

\* We are limiting our discussion to two-dimensional forms presented in the frontal parallel plane.

others. From the initial blur, there gradually emerges a segregated visual object.

The alternative explanation argues that the sensory impulses are structured according to selective principles of organization which do not depend on learning. Such selective principles were first described by Max Wertheimer (80), who referred to them as laws of grouping. Using line and dot patterns, he demonstrated that grouping was not a random arbitrary affair but occurred according to definite principles such as proximity, similarity of color and size, good continuation, etc. Wertheimer described and illustrated these principles with dots and lines but he did not in any way imply that grouping factors operate only with such stimuli. These factors are intended as an explanation for object and form perception in general. Perhaps not all of these factors are necessary to explain organization. Some may, as a matter of fact, prove to be incorrect; nevertheless, organizing factors of this kind seem indispensable for the explanation of form perception. The following illustration demonstrates how these principles would operate to account for the perception of a black circle on a white background.

1. Within the circular contour of the retinal image of the black circle, all points are similar in color as are the points outside the contour (grouping by similarity of color).

2. Within the contour, all points are nearer to each other than to points outside the figure with the exception of points on the contour (grouping by proximity). The following diagram illustrates how this factor would work:

In A one sees two constellations of lines each because of proximity. In B the two constellations are even



FIG. 3

further unified and segregated from each other. In C the grouping is still better and now we have two distinct forms as in everyday life (modified from Köhler [30]).

Thus, similarity and proximity explain why the points within the contour are grouped together and separately from the points outside the contour. As already noted, it is often not realized that if the retinal image is an unstructured mosaic of stimulation, there is no reason why points within a contour should not be seen as belonging with points outside the contour. That they are not is precisely what calls for explanation.

3. The principles of proximity and similarity are insufficient to explain why the circle instead of the surrounding area appears as a shaped entity. To account for this fact, we must invoke another principle of organization first described by Rubin (58). Physically, a contour line serves as a boundary for two areas; it, therefore, can be described as belonging to both. Phenomenally, however, the contour belongs to only one area, giving shape to that area which thereby becomes the figure. The other area remains shapeless and is seen as the ground. This biased belonging of contour must be due to a selective principle—figure-ground organization.<sup>10</sup>

<sup>10</sup> In those cases where conditions are ambiguous, the figure-ground organization is labile and easily reverses itself. The phenomenon of reversible figures in general (i.e., including other types such as the Necker cube and the Schroeder staircase) again shows very clearly that the retinal stimulus does not contain all that is needed for an explanation of the percept.

In the above example, the contour belongs to the black area (because the surrounded region is favored in figure-ground organization), thereby giving rise to the percept of a black circle on a white ground. Most instances of two-dimensional form perception would seem to be accounted for by these laws of grouping (proximity and similarity) and the biased belonging of the contour (figure-ground).

The principles of organization can be considered as purely descriptive generalizations. One can employ these principles without invoking any theory of brain function; nor would any particular type of brain model be demanded. More specifically, the value of the grouping factors as explanatory concepts for form perception does not depend on the physiological theory developed by Köhler in connection with figural aftereffects (34). This theory is designed to deal with the question of how the cortical pattern of excitation is related to the phenomenal size and shape of the percept. By relating the percept to *functional* distance in the cortex (i.e., degree of interaction of current fields) rather than to geometrical distance, and by treating satiation as a changed resistance of the medium to interaction, Köhler was able to account for the facts of figural aftereffects. But even if this theory is correct, it does not eliminate the need for grouping principles, about which it says nothing.<sup>11</sup>

<sup>11</sup> Recently, however, Köhler has suggested that the flow of direct current which he has found to accompany perception may also explain why the figure appears as segregated and distinct from the ground. The current within the cortical correlate of the figure is considered to be highly concentrated and sharply segregated from that of the ground (33). Whether or not this particular idea is correct, it must be admitted that some physiological theory could also adequately explain

In the present paper, we have limited the discussion of form perception to the problem of organization. In doing so we simply assume that the relative position of points to one another in the visual cortex corresponds very closely with the relative position of points in the perceived scene, thus avoiding for the present the question of whether some theory of functional distance is also necessary. Moreover, no attempt is made here to explain why the distortions of the retinal pattern which occur in its cortical projection do not lead to distortions in perception.

An additional problem in form perception arises in connection with whole (Gestalt) qualities and with the fact of transposition (i.e., the phenomenal equivalence of form transposed in size, color, position, etc.). There must be some aspect of organization which underlies the whole quality and which distinguishes one form from another. Underlying the perception of a circle, for example, there must be some characteristic pattern of interaction (one might speculate that it would be symmetrical in some way). This pattern yields the whole quality of circularity, which can also be produced by other patterns of the same form but of a different size or color, since the same characteristic interaction occurs in each case. Although this problem has not been solved, the fact of whole qualities and their transposability represents one of the strongest arguments for the existence of spontaneous processes of organization.

Hebb (23) has outlined an empirical theory of form perception directed toward answering the major

---

grouping. The important consideration is that *some* unlearned law of grouping is necessary, whether stated in purely descriptive terms or in terms of brain events.

arguments of Gestalt theory, many of which are recapitulated in this paper. Space prevents a detailed analysis of Hebb's theory, although much of the evidence he cites is evaluated below. Hebb is not primarily concerned with phenomenal facts in perception but rather with the problem of explaining the *response* to stimulation. Consequently, he discusses only some of the problems considered here.

Hebb grants that a form has "primitive unity" (Hebb's term for figure-ground organization) prior to learning but (if we understand him correctly) this unity does not suffice for form perception. The authors find this point difficult to grasp. Would not the primitive unity of an extended figure (e.g., a straight line) be phenomenally different from that of a compact one (e.g., a solid-color circle)? If so, the admission of primitive unity implies that form perception is not learned. Figure-ground organization means that the contour belongs to the figure, thereby giving it *shape*. Moreover, Hebb does not make clear how the emergence of "cell assemblies" (integrated networks of excitation in the visual areas) changes the phenomenal experience of a form in any way beyond its primitive unity.

Hebb believes that many facts of memory suggest that the memory trace must entail a structural (i.e., physical) change in a specific locus in the brain (a conclusion which he erroneously believes is denied by Gestalt theory) (23, p. 12 ff.). His concept of cell assembly as a particular kind of neural change in a specific locus, characteristic of the stimulus, is intended to explain recognition, i.e., how the same response can be made to a transposed form. Presumably a multiplicity of such assemblies for a given stimulus-pattern are established in all possible positions and to

each the same response is associated. To us, however, transposition suggests that the essential correlate of phenomenal shape is the process to which the stimulus gives rise and not the place in the cortex where it occurs. But this does not imply that the memory trace is unlocalized. Actually, those sympathetic to the organization theory have often speculated that the trace may be localized and recently some evidence for trace localization has been reported (76). They merely stress the fact that, for recognition to occur, a later perceptual process need not occur in the same place as the trace.

Confusion arises when "appearance" is made synonymous with "response."<sup>12</sup> Eventually, of course, a theory is needed to explain how stimulation of different cortical cells can result in the same motor response. The authors believe, however, that premature preoccupation with this problem has led behavioristic psychologists in the wrong direction.<sup>13</sup> It is possible for two similar forms (projected to different loci in the visual areas) to look alike because of similar cortical processes prior to the development of motor responses. Later on, motor development makes possible the association of a specific response (on the human level, the appropriate word) to these similar percepts.

#### LOGICAL DIFFICULTIES INHERENT IN THE EMPIRISTIC VIEW OF FORM PERCEPTION

The statement that the organization of the visual field into shaped regions is learned must mean that at an

<sup>12</sup> The widespread use of the term "perceptual response" is a clear illustration of this identification.

<sup>13</sup> For a lucid discussion of the necessity to deal with phenomenal data in perception, see Allport (2, Chap. 2).

early period in the life history of an organism the visual world does not consist of segregated and unified objects, but appears instead as a mosaic of sense impressions. In some way learning and experience must then transform the sensory data into shaped visual areas. One argument in support of this position is that our visual field contains segregated forms because of previous experience *with those particular forms*. This of course implies that memory traces of previous percepts play a causal role, i.e., they serve to bring about the emergence of those forms when the same stimulus situation occurs later. But how can a memory trace left by an unorganized mass of sensory data create a shaped visual object in the present field? If the initial perceptions consist of amorphous sensations, then how can the memory of such perceptions organize subsequent processes? Instead of trying to explain how the shaped object arises for the first time out of the chaos of sensation, it would seem much simpler to admit some degree of visual segregation resulting from innate organizing processes. One might then say that the influence of past experience must be secondary to spontaneous organization. This logical difficulty inherent in the empiristic theory can be expressed by asking, "How can we learn to see, if we must see in order to learn?"

Empiricists in the past maintained that organization does not first arise in vision but comes about through the sense of touch. By means of tactful exploration of the environment, the child presumably becomes aware of forms and in some way the tactful form causes segregation in the visual field as well. As Köhler has pointed out (32), however, this argument merely transfers the problem of or-

ganization from the visual modality to that of touch. It is still necessary to explain how the discrete tactile sensations can yield an experience of a single object. Moreover, it is difficult to understand how the tactile experience can be transformed into a shaped *visual* object and why there should be such excellent correspondence between the two. Nor is there adequate evidence that touch can yield the precision which we have in visual form perception.

These criticisms also apply to the concept of purposive action as a creative agent of the percept. The transactionalists have not explained how the results of action (which must themselves be perceived) determine the nature of subsequent perception.

A more sophisticated argument for the empiristic theory of form perception might be made by assuming that the principles of grouping are learned (cf. 48, p. 215). This would allow for the transfer of effects of experience to the perception of novel forms; the earlier argument cited above would not, and is, therefore, of limited value. For example, perhaps the child learns that adjacent and similar stimulus elements belong to one object. Even if this is granted, it is still necessary to account for the first emergence of a visual unit. How does the child learn that these stimuli belong together? Does such learning occur because the child sees the object move as a whole when it is manipulated? If so, then "moving together" (Wertheimer's law of common fate) is implicitly accepted as an *unlearned* organizing principle. At some point, the assumption of innate organizing principles must be made in order to explain how learning itself is possible.

Another logical difficulty involved in the effort to explain how specific

past experience modifies subsequent perception relates to the problem of trace *selection*. Even if it were assumed that previous learning has resulted in an organized memory trace for a particular form, this trace cannot exert an influence when the same stimulus is presented again unless this trace and no other is aroused. One way in which the trace of a previous percept could be aroused and thus influence the unorganized sensory impulses would be for the latter to travel to the locus in the nervous system where the relevant trace is "stored." Contact in this way might occur if the successive images of a given object always occurred in the same place on the retina, but this is rarely, if ever, the case. Consequently, as Köhler has argued in elaboration of a point made by Höffding many years ago, appropriate trace arousal must depend on some kind of similarity between the present perceptual process and the trace left by the previous process (31). This means that the present perceptual process must be organized *before* it can communicate with the trace, because only an organized process (i.e., resulting in a definite shape in the case of form perception) can be similar to the trace representing the previously seen form. If the sensory stimuli are unorganized, it is difficult to understand how the proper trace can be selected from the multitude of traces existing in the nervous system. In general, then, past experience cannot exert any influence until the sensory processes themselves are organized.<sup>14</sup>

<sup>14</sup> The same argument arises in connection with experiments purporting to show an influence of motivation on form perception. If a motive is to affect a percept, it would have to do so via memory traces of need-related objects. In one experiment (59), for example,

### THE DISTINCTION BETWEEN PERCEPTION AND RELATED PSYCHOLOGICAL PROCESSES

The term perception has suffered an extension of meaning so broad as to include almost every psychological process. For example, an experiment (7) which obviously deals with recall—the subjects having to reproduce previously seen figures from memory—is widely cited as an experiment in perception. Certain distinctions must be made, not to serve a theoretical bias, but in order to understand the particular process and its relation to other psychological functions.

Perception should, first of all, be differentiated from recognition. Recognition implies a feeling of familiarity—I experience the present object as something I have seen before. The first time the object was seen, however, the perceptual experience occurred without the element of familiarity. In terms of underlying functions, recognition implies that the memory trace of the object is aroused by the present perceptual process; activation of the trace is the basis for the experienced familiarity. By definition, therefore, recognition is dependent upon past experience.

One implication of this distinction is that even if the *same* form were presented repeatedly, the same perceptual experience could conceivably occur each time without recognition.

---

two profiles (one of which has been rewarded in training sessions and the other punished) are presented together to form an ambiguous figure-ground pattern. If the subject is to see the rewarded rather than the punished profile, the memory trace of the former must be the one which has the greater influence. But how can this trace be selected *prior* to the occurrence of figure-ground organization when presumably no shape is as yet seen which is similar to the rewarded or punished face? (For a further discussion of this problem, see references 57 and 75.)

This point may be clarified by reference to an imaginary experiment: *S* views a figure and describes it. Let us assume that the memory trace for this form is destroyed. Later the figure is presented again and *S* is asked to describe it. He is likely to give the same description as he did on the first presentation, even though he is not aware of having seen the figure before. The experience is the same because the stimulus gives rise to the same process in the brain. Recognition represents an additional step—the arousal of the appropriate trace.<sup>16</sup>

As stated above, it seems necessary to assume that trace contact and arousal are mediated by the similarity of the present perceptual process to the trace left by a previous visual experience of the particular object (although there is no explanation available at the present time as to how such a process of trace contact can occur). Even if the present percept is changed or attenuated to some extent, trace contact can still occur as long as there is some formal or structural similarity between the percept and trace. This means that recognition can occur even when material is exposed under unfavorable perceptual conditions (e.g., tachistoscopic presentation, peripheral vision, or dim illumination); moreover, it is reasonable to suppose that it will occur more readily in the case of frequently experienced forms (cf. reference no. 24 and the recent work on the tachistoscopic recognition of words [67]). Recognition of the material does not mean, however, that the percept *qua* form is affected. For

example, a nearsighted person may recognize a friend from a distance but the recognition does not make the percept any clearer; his visual experience is still fuzzy and blurred.<sup>16</sup>

Closely related to the quality of familiarity is the distinctiveness or "identifiability" of certain forms which comes about only with repeated experience. Hebb points out, for example, that, at first, all chimpanzees look alike; with continued observation one begins to recognize individual animals (a similar fact concerning difficulty in distinguishing faces of members of a different racial group from one's own has been mentioned by social psychologists). Frequently, differences among similar objects are not phenomenally registered in initial perceptions; with greater experience, these differences become manifest. This seems to be true, however, only for complex forms. There is no evidence that recurrent observation is necessary in order for a circle and a triangle, for example, to appear as distinct forms. The problem of the discriminability of similar complex patterns requires further investigation,<sup>17</sup> but it should not be confused with the question of form perception *per se*. Past experience is involved in the former case:

<sup>16</sup> A recent experiment by Engel (15) on binocular rivalry between an upright and an inverted face may be another instance where a recognition effect is considered to be a perceptual one. The subjects in this experiment are reported as having seen the upright face more frequently. This result may mean that out of the array of superimposed stimuli, they more readily recognized an upright rather than an inverted face. In our opinion, there is as yet no conclusive evidence that the stimulus elements of the inverted face are suppressed.

<sup>17</sup> Gibson and Gibson (20) have recently performed an interesting experiment to explore this process, which they call "perceptual learning."

<sup>16</sup> The same point applies as well to a *transposed* form. Gestalt psychologists often stressed the recognizability of a transposed structure, such as a melody. But even if not recognized upon repeated hearings, the melody may give rise to a similar experience each time.

repeated perceptions of the form may serve to strengthen memory traces of the details and of the relation of parts to the over-all pattern. These traces provide the basis for an increased awareness both of the internal structure of the form and its difference from similar patterns.

It is also essential to distinguish perception from interpretation. A good deal of the evidence concerning the effects of past experience or motivation on perception actually refers to the process of interpretation. Form perception has been defined as the experience of a segregated object of a certain shape in the visual field; interpretation, on the other hand, refers to the meaning which the visual form has for the subject. Unlike form, meaning is not an outcome of the present stimulus pattern; meaning consists of those qualities and properties acquired by an object through association and learning. On the functional level, meaning derives from the memory traces which are associated with the trace of the visual form itself (e.g., a hammer has meaning because on previous occasions we have seen this particular form used in a certain way; this use is preserved in traces which are associated with the trace of the form percept).<sup>18</sup> This distinction is difficult to make clear because, phenomenally, we perceive

meaningful objects; usually, we do not first experience a pure form percept and then become aware of its meaning. On the level of experience, the meaning is given in the percept, but functionally, two processes must be distinguished.<sup>19</sup>

In some sense modalities, the distinction we are making does frequently appear in experience; e.g., I hear a sound and then try to identify it—the cry of a baby or meow of a cat. Even in vision the separation of processes may be experienced. For example, a nonsense form is seen as a segregated unit of a certain shape; nevertheless, it may have little or no meaning, and one may strive to interpret it. (Moreover, after gaining meaning, the form itself does not change in my experience. At first,  $\text{J}$  was a meaningless shape. Although I now see it as an eighth-note the visual form has not altered in any way.) The separation of processes may be more evident in the child's experience than in the adult's. It seems probable that the child sees objects before he has any concept of their meaning.

The separation of the perceptual from the interpretive process is not an arbitrary matter of definition; on the contrary, it is necessary to make this distinction in order to account for the nature of our experience. It is important also to keep this distinction in mind when evaluating experimental studies of perceptual problems. For example, if we should want to describe correctly the initial perceptions of congenitally blind subjects whose vision had been restored,

<sup>18</sup> There has been some confusion concerning the treatment of meaning in Gestalt psychology. Apparently, some of the earlier writings of Gestalt psychologists created the impression that meaning was thought to be given directly in the present percept. (The contact between Gestalt psychology and philosophical phenomenology may have contributed to this impression.) It seems to the present writers that meaning must be explained in terms of associated traces or trace systems, and is, therefore, derived from past experience. Köhler has stated this position very clearly (32, p. 138 ff.).

<sup>19</sup> The same is true about the distinction between perception and recognition. Phenomenally, familiarity is in the object; functionally, one must assume that familiarity derives from trace reference after the perceptual process occurs.

we should not confuse their failure to identify objects with an inability to perceive objects as segregated units. Much of the data collected by von Senden (61) is vitiated because the investigators did not clearly distinguish the two functions. We may be distorting the experience of a hungry subject who describes an ambiguous shape as a steak if we conclude that the hunger drive has affected his perception (cf. 39). Possibly, he sees the same form as does the nonhungry subject, but interprets it differently. (If asked to copy the form, both subjects might make fairly identical drawings.)

The Rorschach test, insofar as it is concerned with the ways in which shapes are described (leaving aside the color, shading, and other aspects), is primarily a test of interpretation. Many meanings can be ascribed to the blot as a whole or to a particular part.

#### EXPERIMENTAL EVIDENCE

The major portion of the following section will be devoted to a critical analysis of some representative studies dealing with the question of whether form perception is innately determined. To begin with, some evidence relating to the determinants of other perceptual processes will be briefly cited but there is no intention of making a comprehensive coverage of the literature bearing on this issue. Many studies are inconclusive because no attempt was made to control the effects of previous experience.

*Visual direction.* Schlodtman (60) showed that congenitally blind subjects localized the direction of pressure phosphenes in the same way as do normally sighted subjects. More recently, Hess (26) has confirmed earlier findings (e.g., 4, 11) that chicks peck in directions innately de-

termined by retinal locus. Sperry's experiments (68) provide further support of the thesis that visual direction is unlearned.<sup>20</sup>

*Visual constancies.* The constancies—size, color, and brightness—have been shown to exist in various animal species (cf. reference no. 40 for a summary of the literature). Size constancy, for example, has been demonstrated in a three-month-old chicken (22) and in eleven-month-old infants (17). Although these studies are not crucial for the issue of innateness, they would appear to conflict with naive empiristic views which account for constancy on the basis of knowledge or unconscious inference.

*Depth perception.* There seems to be little unequivocal evidence relating to the problem of distance or depth perception. Lashley and Russell (36) concluded that visual depth was innately determined in rats, and Hess succeeded in showing that chickens with no previous visual experience (or with prior alternating monocular vision) utilized binocular depth cues (26).

*Visual reflexes.* Observations on infants reveal that some visual-motor coordinations, such as eyelid responses to intense light, pursuit movements, and fixation are present at birth, or soon after (12, 50). These data, however, are not entirely relevant to the study of visual experience, since they may simply represent reflex responses to stimulation by light without being accompanied by the perception of direction, color, form, or depth.

*Form.* Two major experimental approaches have been employed to determine the effects of past experience on form perception. The first group of studies we shall discuss attempts

<sup>20</sup> Caution is necessary in generalizing the results of animal experimentation, in perception as well as other areas, to the human level.

to show how experimentally created familiarity with specific forms affects subsequent perception; the second group attacks the problem more directly by studying the consequences of the deprivation of normal visual stimulation soon after birth on later perceptual development. Included in the latter approach are the observations on congenitally blind humans who gained vision in later life.

The classic experiment in the first group is the investigation by Gottschaldt (21). Gottschaldt wanted to show that a novel geometrical figure will be seen in accordance with the laws of grouping rather than past experience. He reasoned that if form perception were determined exclusively by experiential factors, a complex figure *b*, containing a simple form *a* which has been seen very frequently in the past, should be perceived as the familiar unit *a* plus other parts. Gottschaldt designed some simple outline figures which were presented repeatedly to subjects for memorization. Later, complex figures in which the *a* figures were embedded were shown and the subjects were instructed to describe them. It was found that only in a negligible number of cases was *b* spontaneously described as the *a* figure and additional lines. Despite its great familiarity at the time of the test, *a* was not seen.

Gottschaldt's experiment has been criticized on the ground that it merely shows that familiar units can be camouflaged by embedding them in larger contexts. This criticism misses the point completely because it fails to see the necessity for explaining why the physically present figure is phenomenally absent. The camouflage is successful because of the victory of grouping factors over past experience. Not just any additional lines will successfully camouflage the *a*

figure but only those which, because of the laws of grouping, produce new and compelling organizations. A few well-placed lines will achieve this effect, whereas a complex array of lines may not succeed in camouflaging the *a* figure (cf. 32, p. 193 ff.). Good continuation is probably the strongest factor in Gottschaldt's figures. Camouflage in nature, which of course involves additional factors (e.g., countershading, similarity of color, etc.), also demonstrates that familiar objects will not be readily perceived when they are in certain environmental backgrounds (10, 42).

According to Hebb, Gottschaldt's conclusion "is valid only if the total figure is an unanalyzable whole, which it surely is not" (23, p. 24). One *b* diagram, for example, contained two parallelograms and a set of lines forming a Z. Hebb implies that the presence of these familiar units in the *b* figure explains why the *a* figure was not seen. It is possible, however, to embed the *a* figure in a *b* diagram which contains familiar

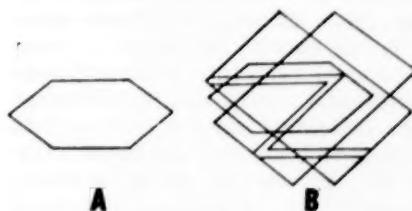


FIG. 4. A. ONE OF GOTTSCHALDT'S SIMPLE FIGURES; B. A MODIFIED VERSION OF GOTTSCHALDT'S COMPLEX FIGURE CONTAINING "A," TO WHICH HEBB REFERS

parts and still the simple form will stand out. It is the structure of the total figure which is crucial and not the familiarity of any of its parts. Moreover, even if the subject distinguishes such familiar parts in the complex diagram, the question still

remains whether this is due to familiarity or to structure.<sup>21</sup>

According to some critics (28), Gottschaldt's thesis cannot be accepted because it has been refuted by later investigators. The experiment by Djang (13) is often cited in this connection.

Djang's results show a strong effect of past experience. Simple figures which had previously been learned were found in complex forms twenty times more frequently than those not seen before. (All figures were composed of dotted lines; the subjects had to learn to draw the figure correctly from immediate memory and to associate a nonsense name to each figure. The task was described as one of learning and memory.)

Careful examination of the conditions of Djang's study makes it apparent that her results do not invalidate or even challenge Gottschaldt's conclusions. Of special significance are the following aspects of the experiment.

1. The instructions encouraged the subject to break up the complex figure into individual sections or units. "Try at once to reproduce . . . what you have seen . . . Indicate by the additional use of the yellow pencil the individual units or sections into which you split up the figure" (13, p. 34). Evidence for seeing the simple figure in the complex one was based on the units which the subjects encircled. The complex figure contained many subunits so that, in addition to seeing the figure as a whole, the subject with this set might be expected to see now one part and now another as

a relatively separate entity. When the subject recognizes a unit as one he has previously seen, he is likely to encircle it. This would facilitate the learning of the complex figure, since the subunit represents a substantial portion of the total figure. That such a set was important is shown by the author's remark that "success in finding the simple figure in the camouflage seems to bear a relation to the amount of interest displayed" (p. 47). Enthusiastic subjects were the most successful. Djang does show that her data cannot be explained merely as a result of a set to look for familiar units. But a set to break up the figures into parts is an important condition for the effect.

2. Unlike Gottschaldt's simple figures which were absorbed into the larger structure, many of Djang's are isolable subunits because their contours are not destroyed by good continuation. Since the camouflaging effect of Gottschaldt's figures is an essential feature of his design, one may question the construction of Djang's figures. Her results which are based on the use of these figures can, therefore, in no way affect the validity of Gottschaldt's conclusions.

3. Even without an analytical set, the subjects in this experiment might take note of a simple form in a complex one because they *recognize* it; this is true because of the point made in paragraph 2 above. In Gottschaldt's experiment, the *a* figure was not recognized because it was not seen. In Djang's study, however, both the simple unit which had not been seen in prior exposures and the one which is recognized may have been perceived (if only briefly) in the complex figure with equal frequency; but if the subjects had not seen the simple unit before, there would be no special reason to notice it. In other words,

<sup>21</sup> Hebb also points out that if one looks for the simple form, one can find it. Here, of course, he is referring to the problem of the influence of attention or set and we agree that no psychological theory has as yet provided a satisfactory answer to this problem.

Djang has not proved that there is a difference in frequency of *perception* of the simple form but only that there is a difference in the *utilization* of this form based on recognition.

4. The fact that some masked figures were more easily found than others cannot be explained by previous experience (which was equal for all simple forms) but must be understood on the basis of factors of organization. Those masked figures whose contours do not continue into the other lines of the complex form should be readily seen; on the other hand, the use of good continuation should lead to fewer successes. Figures LAJ, ZIF, and GIW have the least number of successes—in these figures the contour of the simple form is to some extent continued into the larger structure. The influence of past experience is greatest with figures XEH, QOW, POQ, KOJ—and these are the simple forms which are easily segregated from the larger form. It is possible to take one of Djang's figures and, by strengthening organizational factors, make it difficult for the familiar simple figure to emerge (see Fig. 5).

This shows conclusively that it is not the use of dot figures which distinguishes Djang's experiment from that of Gottschaldt. It is the construction of the dot figures, together with her procedure, which made her results possible. As a matter of fact, this experiment supports Gottschaldt's contention that strong structural factors overcome the effects of familiarity.

Braly (5) attempted to show that the perception of polygonal, dot figures is influenced by the kind of figures shown earlier. The test slides, however, contain several of these dot figures and it is impossible to see all of them clearly in the very brief ex-

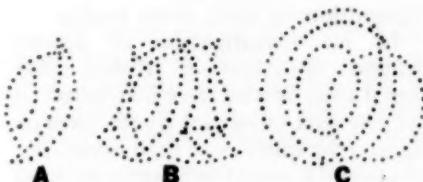


FIG. 5. A. ONE OF DJANG'S SIMPLE FIGURES (QEW). B. DJANG'S COMPLEX FIGURE CONTAINING QEW (KOJ). C. A MODIFIED VERSION OF KOJ CONTAINING QEW

posure time. The experiment demonstrates only that given a set to perceive a certain form and subsequently given inadequate perceptual conditions, Ss will tend to guess in accordance with that set.

Henle (24) posed the question whether a familiar form would be more readily perceived than an unfamiliar one when structure is held constant. A series of letters and numbers and their mirror reversals (together with obverse and reverse nonsense forms) was exposed peripherally or tachistoscopically. The results, based on the Ss' reproduction of the forms, show that the obverse letters were reproduced correctly more frequently than the reverse letters. Does the experiment demonstrate an influence of familiarity on *perception*? Perhaps the obverse letters are not more clearly perceived than their mirror reversals, but are more easily recognized under difficult perceptual conditions because of their familiarity. Once recognized, a familiar letter is easy to draw. The reverse letter would probably be seen as a nonsense figure, and, consequently, the subject is faced with the added difficulty of remembering its inadequately perceived shape in order to draw it a few moments later. Following the analysis given above, we would argue that the presence of stronger trace systems for obverse letters allows

recognition to occur more readily.

In his investigation of figure-ground organization, Rubin (58) found that Ss who were instructed to see one part of an ambiguous form as figure (the other part then appearing as ground), would subsequently tend to see the same part as figure. If Rubin's results were valid, they would certainly provide evidence that memory traces can organize perceptual processes. Recently, however, a careful repetition of Rubin's experiment by Rock and Kremen (57) failed to demonstrate this effect.

Leeper (38) found that subjects would generally see Street figures as meaningless collections of fragments upon first presentation. After a brief period of observation (sometimes accompanied by verbal hints from the experimenter), the figures were reorganized and perceived as meaningful objects. Several weeks later, when the same Street figures were exposed tachistoscopically, they were immediately recognized in their meaningful form. Leeper's experiment does show a past experience effect, and thus seems to contradict the logical argument that traces cannot influence perceptual processes until the latter are organized. This specific problem will be discussed below.

We turn now to a consideration of the more direct kind of evidence. It would appear that a crucial test of the empiristic and organization theories could be provided by a "deprivation" experiment in which no opportunity to learn form perception through visual experience is permitted during the organism's early life.

On the human level, the data consist of observations made on cases of congenital blindness (due to cataracts) to whom vision was restored in later life by surgical operation. The literature on such cases has been an-

alyzed by von Senden (61) and his study is often cited in support of the empiristic theory. According to Hebb, for example, these patients could not immediately distinguish forms after vision was gained; a long, gradual, learning process was necessary to enable the patients to perceive. There are, however, serious deficiencies in this evidence (cf. Michael Wertheimer [81], who describes some of the flaws and, in addition, observes that von Senden has often been cited erroneously). The conditions and the exact time after operation of the observations were not adequately described; the extent of vision present before operation varied from case to case; some of the cases were young children whose reports are difficult to evaluate. Moreover, the patients, after operation, were faced with a strange new world and often the investigator (usually the surgeon) did not know what questions to ask, or what tests to perform, in order to elicit the subject's experience. In one case, for example, the patient "had great difficulty in describing her sensations in such a way as to convey any clear conception of them to another" (37, p. 148). Much of this evidence, therefore, is inconclusive.

With respect to form perception, it appears that no distinction was made in these studies between perceptual and interpretive processes. In the eighteenth and nineteenth centuries (when most of the cases studied by von Senden occurred) the problem was posed by investigators in the following way: Would a blind person, who can distinguish a sphere from a cube by means of touch, be able to identify these forms visually when seen for the first time? Observations of these newly sighted patients seemed to show that they could not. There is,

however, no reason to expect such a result. The patient might see the sphere and cube as different forms but would not know their appropriate names until permitted the use of touch. Moreover, even if told which was which, he would have to remember this information, so that further learning would be required for correct identification, although not for perceptual discrimination.

It is clear from some of these cases that the visual field of the patient was not an undifferentiated blur but did consist of forms and shapes which could be perceived but, of course, not named. Frequently, the case report describes the patient looking at something and asking "what is that?" One intelligent patient, as a matter of fact, was able to identify a ball as round and a toy brick as square upon first presentation (37). In a more recent case (16), the report also suggests that the patient could see objects but was not able to identify them. The observations on newly sighted patients, therefore, in no way lend support to an empiristic theory of form perception.

More carefully controlled observations are, of course, possible with animals. In recent years, there have been a number of investigations of the effects of early visual deprivation upon subsequent perceptual behavior.

Siegel (63), in a carefully designed experiment, raised a group of ring doves with plastic head covers which permitted light stimulation but no pattern vision. The hoods were put into place soon after the birds were hatched and were worn for a period of from eight to twelve weeks. A control group of birds was raised in a normal visual environment. At the end of this pretraining period, window openings were cut in the hoods

of the experimental doves. Both groups were now trained on a visual form discrimination—jumping to a triangle vs. a circle. Thirty trials were given on each training day. The criterion for successful discrimination was nine out of ten consecutive jumps to the positive stimulus on any given day of the training series.

Siegel found that the hood-reared birds required an average of 126.8 trials to reach the criterion, while the controls required 77.7 trials; the difference between the groups was statistically significant. These results, according to Siegel, tend to verify theories which stress the crucial role of past experience in perception. Actually, they may be interpreted as furnishing cogent evidence for a non-learning position. If form perception must be learned, it is very surprising that after eight to twelve weeks of homogeneous light stimulation the experimental birds required only 49.1 additional trials for correct performance (one and two-thirds additional training days). Moreover, Siegel's published report gives only group data; the individual performance records (65) show that one or two hood-reared birds were able to respond correctly very soon after unhooding. For example, experimental bird no. 13 required only 58 trials to reach the criterion; this performance is better than that of eleven out of twelve controls and practically on a par with the best of the controls (no. 26) who required 50 trials. It must also be pointed out that the group difference obtained by Siegel refers to the arbitrary criterion for success of nine out of ten correct trials. But eight out of ten correct jumps (or even seven of ten) for two or more blocks of ten in succession is certainly above chance performance. We do not know whether a significant

difference would be obtained for this criterion.

It must be remembered that form discrimination is not a test of perception alone; cognitive factors are also involved. It is possible that the animal perceptually distinguishes the triangle and circle but requires training in order to learn that response to one stimulus is followed by reward. It might take a human subject two or three trials before he realizes that he must respond to a triangle and not to a circle. Would anyone argue from this fact that on these first few trials the subject did not see two different forms? It is not surprising that an animal deprived of visual form experience for the first several months of life would show some retardation in solving a discrimination problem. (We are not referring to any difficulty in motor performance; no objection to the experiment on such grounds seems justified since Siegel took the precaution of forcing the animals to jump from a platform a total of four hundred times before the hoods were removed.)

An interesting experiment by Miller (44) is relevant for the interpretation of visual deprivation studies such as those of Siegel and Riesen. Miller's hypothesis was that the first visual experience of an animal raised in darkness may create "a negative disturbance effect which inhibits instantaneous utilization of the new cues even though perception may be immediate and accurate" (44, p. 224). He raised a group of rats in a light-proof cage and a control group in a normal visual environment. At sixty-five days of age both groups were trained to run an obstacle course *in the dark*. (After each trial the experimental animals were returned to their dark cages.) When both groups had learned to perform rapidly, the

lights were turned on for the run. The controls did not seem to be affected by the light. The experimental rats, for whom this was the first visual experience, showed a significantly longer mean running time, and an increase of inter-rat variability (individual running times ranged from six to twenty-nine seconds). The experiment shows that performance on a task already learned on the basis of other sensory cues may be disturbed by the new visual experience. Therefore it is probable that such a disturbance would be present in the learning of new tasks, as in Siegel's experiment. In addition, Miller's results point up the importance of taking individual differences into account in studies of this kind.

The *earlier* work of Riesen (51, 52), in which chimpanzees were raised for a long period of time in total darkness, requires only brief mention for the purposes of this paper. The visual defects shown by these animals may have been due to optic atrophy rather than to the lack of opportunity for learning (79). The more recent investigations (53), on the other hand, are very important for the problem of learning in perception.

In the revised procedure, chimpanzees were placed in a dark room five days after birth. For 90 minutes each day, the animal's head was enclosed in a Plexiglas dome which permitted stimulation by diffused light. This procedure was continued until the animal was seven and one-half months old when gradually, over a period of ten days, it was given more and more light (increased illumination of the room). In addition to observing the animal's behavior in relation to visual objects, the following experiment was performed. Training of an avoidance response was be-

gun; twice a day, a shock plaque (a disk painted with vertical yellow and black stripes) was held in front of the animal and brought slowly toward him until an electrode made contact with his face and delivered a shock. When an avoidance response had been established, discrimination training was started. The shock plaque was shown, followed by shock if the animal did not make an avoidance response. Four other plaques were always followed by the food bottle. These "positive" disks differed from the negative stimulus in either one of the following characteristics: size, color, shape, and direction of stripes. Complete data are reported for only one such animal, Chow, and for two other animals—Faik, reared normally, and Lad, reared like Chow except that he received 90 minutes a day of patterned light stimulation.

Riesen reports that Chow and Kora (another chimpanzee reared in the same way as Chow) evinced difficulty in learning to recognize objects such as the food bottle. In the experiment, Chow showed delay in avoiding the shock plaque as compared to the normal control. Also his performance for the discrimination series as a whole was inferior to that of Faik or Lad. There are, however, discrepancies in the data which make interpretation difficult. For example, although Chow made many more errors than Faik before reaching the criterion for the shape discrimination, he was superior in learning the discrimination between the horizontal and vertical stripes. Certainly discrimination of the direction of stripes shows some degree of form perception. Chow had the greatest difficulty in discriminating size and shape and little difficulty with color as well as direction of stripes. The difficulty may be a cognitive one—i.e., per-

haps it was difficult for Chow to abstract the size and shape characteristics from the more striking surface features of the plaque. The fact that Chow made more errors than Faik on the discrimination of *size* supports this interpretation; empiristic theory does not imply that the perception of size differences in objects of the same shape presented at the same distance must be learned. Another finding which is hard to understand is the fact that Lad, who had only 90 minutes a day of pattern vision for the first seven months of life, made fewer errors to all positive plaques than Faik, the normally reared animal. Yet Lad had many more failures than either Faik or Chow in reaction to the shock plaque. One animal, Mita, reared like Lad, but restricted in a supine position in a holder, apparently also had difficulty in learning to discriminate the bottle from other objects. This fact is not easy to explain.

It is also worth mentioning that for a long time after being placed in a normal environment, chimpanzees who had been reared without pattern vision had difficulty with pursuit of moving objects and with binocular fixation; they also manifested spontaneous nystagmus. Although Riesen's results cannot be fully accounted for by the presence of these impairments, it is plausible that such visual anomalies contributed to difficulty in clearly perceiving a unified and stable world of objects.

We do not wish to minimize the importance and interest of these studies. It is essential, however, to recognize the problems involved in the interpretation of the data. It is clear that no definite conclusions can be reached on the basis of studies employing so few subjects; the discrepancies mentioned above may simply

represent individual differences. Nor do we know what are the effects of a restricted sensory environment on the cognitive maturation of the animal—there is evidence that animals reared under such conditions manifest some impairment of intelligence (70). Furthermore, the results of experiments in which subjects are raised in an abnormal or restricted environment lend themselves to two different interpretations. If a particular function or capacity does not appear or is retarded, this may mean either that learning is necessary or that the experimental conditions have disrupted the normal maturation of a function which may be innate. For example, Nissen, Chow, and Semmes (47) have shown that a chimpanzee who had been reared with little opportunity for tactual experience (his arms and legs were encased in cardboard tubes) was unable to solve a problem requiring tactual discrimination of two widely separated stimulus points. But, the "neonate chimpanzee responds differentially according to the location on his body of a tactual stimulus. Usually the principal movement is near the region of stimulation" (47, p. 494). Similarly, if a pain stimulus is applied to the cheek of a human infant, the infant's hand is brought to the cheek near the point stimulated (62). These facts suggest the possibility that some degree of tactual discrimination is innate and that the restricted experimental conditions have disturbed or prevented normal development so that the chimpanzee is unable to solve the discrimination problem. In spite of these reservations, however, we have no doubt that this type of experiment has brought us closer to a crucial test of the two theories of form perception discussed in this paper.

The effects of the deprivation of pattern vision on interocular transfer have been recently investigated in birds (64), cats (54, 55), and chimpanzees (9). It has been found that if, in rearing, both eyes have been exposed to patterned light (either simultaneously or alternately), the animal later trained monocularly on a visual discrimination problem transfers immediately to the untrained eye. If the animal is reared in darkness and then given diffuse light to one eye and patterned light to the other (or reared with both eyes stimulated by diffuse light), there is no immediate transfer of the discrimination to the untrained eye, regardless of which eye is used in the training; the discriminations are re-learned, however, with considerable savings. These results, while highly significant and quite surprising, are not relevant to the *perception* of form. There is no evidence that the animals could not see forms when the eye which had been given diffuse light was exposed for the transfer tests.<sup>22</sup> In fact, some of the data permit the inference that the lack of immediate transfer could not have been due to any difficulty in perceiving with this eye. First of all, even when an animal was trained with the eye which previously had been stimulated only by diffuse light, there was no transfer to the other eye, which had received patterned light (9). Secondly, an animal trained on three problems in succession failed in each case to transfer to the untrained eye (9). This animal re-learned the first problem with the untrained (dif-

<sup>22</sup> Riesen et al. (54) report that when the diffuse-light eye was first exposed, the cats bumped into objects and moved about quite slowly. But similar behavior occurred when the previously trained eye was re-exposed. Nissen et al. (9) do not report the initial behavior of the chimpanzees in their experiment.

fuse-light) eye so that perception must have been adequate when the second discrimination was begun. Furthermore, it would be difficult to argue that color differences are not perceived with the diffuse-light eye, yet color and brightness discrimination problems show the same effects as discriminations of form (9, 54). These experiments, therefore, relate to the problem of recognition (i.e., accessibility to the trace) and reveal a limitation of the process of trace contact by similarity.<sup>23</sup> We would argue, applying the analysis given earlier (p. 280), that the stimulus seen with the untrained eye has the same appearance for the subject but it is not recognized as one which leads to reward. Why trace contact from one eye to the other does not occur so readily only in the case where both eyes have not had patterned light is, of course, a puzzling problem.

#### WHAT PAST EXPERIENCE CONTRIBUTES TO PERCEPTION

Every theory must grant that some aspects of perception are spontaneous reactions to the stimulus situation; no one, for example, has argued that when the retina is stimulated by light of wavelength 700 mu., learning is required before the color red can be seen. But both logical analysis and empirical evidence support the con-

<sup>23</sup> The fact that some perceptual phenomena diminish in strength or magnitude when transferred to the previously unstimulated eye reveals a similar limitation of interaction. If a rotating spiral is viewed with one eye, the negative aftereffect is greater with the same eye than with the other. Similarly, Gibson (18) found that the aftereffect of the inspection of a curved line is stronger with the eye used during the inspection period than with the eye which had not been stimulated. The influence of past experience on the perception of a wire cube is greater with the eye previously used (1).

clusion that much more than color experience is immediately "given" as a result of innate organizing factors. Specifically, the organization of the visual field into shaped areas is not an outcome of learning—past experience cannot carve visual form out of initially formless perception. Other phenomena in perception considered to be innately determined were referred to earlier (p. 273).

But this does not imply that perception is not affected by past experience. On the contrary, it is only when some degree of innate organization is granted that the effects of learning can be more clearly understood.

1. The role of past experience in lending familiarity, ease of recognition, and discriminability, as well as meaning, has already been discussed. Perceptual experience is greatly enriched by the addition of these aspects; in the light of the distinctions made above, however, form perception as such is not affected.

2. In some cases, a memory trace can reorganize or modify a percept. An experiment by Wallach, O'Connell, & Neisser (78) demonstrates that a memory trace can impart three-dimensionality to a figure which at first was seen as two-dimensional. Wallach presented to a control group the shadow of a wire figure which was described by the subjects as flat. In the presentation to the experimental group, the wire figure was rotated, giving rise to a constantly distorting pattern on the shadow screen. The figure was now seen as three-dimensional (kinetic depth effect). Sometime later, when the same figure was presented in a stationary position (where it had been seen as two-dimensional by the control group), it was described by the experimental subjects as three-dimensional. Certain controls ensured that

the effect was a perceptual and not a cognitive one.<sup>24</sup>

To explain these results in terms of underlying functions we must assume the following: As a result of the moving presentation, a memory trace is left of a three-dimensional form. In the later stationary presentation, contact with this "three-dimensional" trace occurs on the basis of similarity of form and three-dimensionality is thereby imparted to the perceptual experience.

The experiment suggests the possibility that many of the purely figural cues to three-dimensionality (perspective, overlay, specific patterns such as a trapezoid giving rise to the percept of a rectangle at a slant, etc.) may be learned. Although empiricists have assumed such cues to be learned, they have never offered a plausible explanation as to how the learning takes place. In terms of the above hypothesis, the following explanation becomes possible.

Unlearned depth perception occurs on the basis of certain cues such as retinal disparity or the kinetic depth effect, and leaves *visual* traces. These visual traces can impart three-dimensionality to new figures, which otherwise would be perceived as two-dimensional (e.g., the Necker cube, perspective drawings, etc.). This assumption obviates reference to touch or purposive behavior as the source of the learned depth experience. The evocation of past experience effects occurs only when relevant traces are selected by the presence of a stimulus with some similarity to the previous

three-dimensional percepts. By accounting for the initial depth perceptions and for the arousal of the appropriate traces, this hypothesis overcomes the logical difficulties in empiristic theories.

One might explain Leeper's experiment (38) in a similar way. On initial presentation the Street figure may be experienced as a jumble of fragments. After a period of inspection, the figure may suddenly look different—it is now recognized as a meaningful object. How this initial recognition occurs is not clear but in functional terms we may assume that the memory trace of the meaningful object is aroused in some way by the Street figure, and that this trace changes the phenomenal appearance of the figure. It will be recalled that when the Street figures were re-exposed several weeks later, they were instantly seen in their meaningful form. How is this effect to be explained? Following Wallach's reasoning, we may assume that the first presentation leaves behind two traces—one corresponding to the perception of the figure as an unorganized collection of fragments, and associated with it a trace of the figure in its meaningful form. Meaningful re-recognition of the figure means that the second trace is aroused. In accordance with the logical argument stated earlier, arousal of the first trace must occur and only then can the associated trace be activated in order to restructure the percept.<sup>25</sup>

3. Past experience within the experimental situation or experimental

<sup>24</sup> Informal repetition of this experiment at the laboratory of the New School for Social Research has failed to confirm that the effect is as easily obtained as the original report suggests. But even if such a memory effect occurs only occasionally it remains of great importance. In the present discussion, the experiment is cited primarily to illustrate how past experience might modify organization.

<sup>25</sup> Perhaps a similar process (i.e., the modification of a percept by a trace which is aroused by some kind of partial similarity with the still incompletely organized stimulus) could explain the selective influence of previous experience in ambiguous situations. The factual basis for this type of effect, however, is still unclear (cf. 27, 56, 57, 58, 59, 66).

instructions may produce a set which in turn influences the perceptual outcome. As noted above, there is as yet no explanation for the action of a set or attitude in modifying perception.

4. Prior experience may change the neural medium so that subsequent percepts are modified. This is not an effect of past experience in the usual sense because it is not specific to the contents of subsequent perceptions; it affects more or less indiscriminately stimuli which impinge at a later time in a specific region. Examples of this category might include: adaptations of various kinds, figural aftereffects and the negative aftereffect of movement. Evidence concerning the effects of long range adaptation to unusual stimulus conditions has recently appeared (25).

5. The studies of Ivo Kohler (29), however, suggest that adjustment to prism-produced distortions or to chromatic lenses cannot be understood merely in terms of local adaptations since the effects are dependent upon eye position. Thus, for example, Ss wearing glasses, each lens of which consisted of a blue left half and yellow right half, in time adapted so that when the eyes were turned to the left the scene appeared less blue, and when turned to the right, less yellow, than at first. Furthermore, when the glasses were removed, Ss reported aftereffects which are also dependent on eye position. With eyes to the left, the scene looked yellowish; with eyes to the right, it looked bluish. Similar adaptations and "situational aftereffects" occur with respect to distortions caused by the wearing of prisms. It is difficult to assess the full significance of these recently published findings, but clearly an important effect of past experience on perception seems to have been demonstrated.

6. Past experience may have an *indirect* effect by determining conditions which make other processes possible, although these processes themselves are not the results of experiential factors. The apparent oscillation of an objectively rotating trapezoidal window (3) may be understood in this way. If the window is seen as rectangular, other perceptual effects must follow. The perceived rectangularity itself may be due to previous experience<sup>26</sup> (an assumption which could be challenged by those who accept the principle of Prägnanz). Seen as a rectangle, the window cannot come into the frontal-parallel plane and must, therefore, be perceived to oscillate through an angle of less than 180°.

Another possible example of an indirect effect is furnished by the situation where those cues to distance which may be learned give rise to size constancy, which may be innately determined.

#### CONCLUSION

One can hardly take a dogmatic position in an area where, as yet, there exists so little decisive experimentation. Nevertheless, it is important to determine the status of a scientific theory in relation to present knowledge. On the basis of logical analysis and an examination of relevant evidence, we have argued for the thesis that various aspects of the phenomenal world and, in particular, the segregation and shape of visual forms are given by innate organizing processes. Percepts may be modified and enriched by experiential factors but the effects of such factors presup-

\* Explanation of the apparent rectangularity in terms of visual traces makes more concrete the hypothesis which the transactionalists imply by such terms as "assumptions," "prognostic directives for future action," etc.

pose the prior existence of visual forms.

If the thesis defended in this paper is correct, perceptual organization must occur *before* experience (or personality factors which depend on experience, such as need, purpose, and value) can exert any influence. According to holistic concepts, currently so popular, psychological functions cannot be separated. But it is the relative independence of the perceptual organizing processes which

makes possible an adequate phenomenal representation of the external world. Despite changing motives and emotions, phenomenal color, form, and space remain remarkably stable and generally correspond to the objective situation. Such correspondence is, of course, necessary for successful adaptation to the environment and the innate neural processes which yield this correspondence must themselves represent the product of adaptive evolution.

#### REFERENCES

- ADAMS, PAULINE A. The effect of past experience on the perspective reversal of a tridimensional figure. *Amer. J. Psychol.*, 1954, **67**, 708-710.
- ALLPORT, F. H. *Theories of perception and the concept of structure*. New York: Wiley, 1955.
- AMES, A. Visual perception and the rotating trapezoidal window. *Psychol. Monogr.*, 1951, **65**, No. 7 (Whole No. 324).
- BIRD, C. The effect of maturation upon the pecking instinct of chicks. *Ped. Sem.*, 1926, **33**, 212-234.
- BRALY, K. W. The influence of past experience in visual perception. *J. exp. Psychol.*, 1933, **16**, 613-643.
- BROWN, J. F. The visual perception of velocity. *Psychol. Forsch.*, 1931, **14**, 199-232.
- CARMICHAEL, L., HOGAN, H. P., & WALTER, A. A. An experimental study of the effect of language on the reproduction of visually perceived form. *J. exp. Psychol.*, 1932, **15**, 73-86.
- CARR, H. A. *Introduction to space perception*. New York: Longmans, Green, 1935.
- CHOW, K. L., & NISSEN, H. W. Interocular transfer of learning in visually naive and experienced infant chimpanzees. *J. comp. physiol. Psychol.*, 1955, **48**, 229-237.
- COTT, H. B. *Adaptive coloration in animals*. London: Methuen and Co., 1940.
- CRUZE, W. W. Maturation and learning in chicks. *J. comp. Psychol.*, 1935, **19**, 371-409.
- DENNIS, W. Is the newborn infant's repertoire learned or instinctive? *Psychol. Rev.*, 1943, **50**, 330-337.
- DJANG, S. The role of past experience in the visual apprehension of masked forms. *J. exp. Psychol.*, 1937, **20**, 29-59.
- DUNCKER, K. Über induzierte Bewegung. *Psychol. Forsch.*, 1929, **12**, 180-259.
- ENGEL, E. The role of content in binocular resolution. *Amer. J. Psychol.*, 1956, **69**, 87-91.
- FISHER, J. H. Vision learning after successful operation at age 6. *Ophthal. Rev.*, 1914, **33**, 161-165.
- FRANK, H. Untersuchungen über Sehgrößenkonstanz bei Kindern. *Psychol. Forsch.*, 1926, **7**, 137-145.
- GIBSON, J. J. Adaptation, after-effect and contrast in the perception of curved lines. *J. exp. Psychol.*, 1933, **16**, 1-31.
- GIBSON, J. J. *The perception of the visual world*. Boston: Houghton Mifflin, 1950.
- GIBSON, J. J., & GIBSON, ELEANOR J. Perceptual learning: differentiation or enrichment? *Psychol. Rev.*, 1955, **62**, 32-41.
- GOTTSCHALDT, K. Über den Einfluss der Erfahrung auf die Wahrnehmung von Figuren. I. *Psychol. Forsch.*, 1926, **8**, 261-317.
- GÖTZ, W. Experimentelle Untersuchungen zum Problem der Sehgrößenkonstanz beim Haushuhn. *Z. Psychol.*, 1926, **99**, 247-260.
- HEBB, D. O. *The organization of behavior*. New York: Wiley, 1949.
- HENLE, M. An experimental investigation of past experience as a determinant of visual form perception. *J. exp. Psychol.*, 1942, **30**, 1-22.
- HERON, W., & DOANE, B. K., & SCOTT, T. H. Visual disturbances after prolonged perceptual isolation. *Canad. J. Psychol.*, 1956, **10**, 13-18.

26. HESS, E. H. Space perception in the chick. *Sci. Amer.*, 1956, **195**, 71-80.

27. JACKSON, D. N. A further examination of the role of autism in a visual figure-ground relationship. *J. Psychol.*, 1954, **38**, 339-357.

28. KILPATRICK, F. P. (Ed.) *Human behavior from the transactional point of view*. Hanover, N. H.: Inst. for Associated Research, 1952.

29. KOHLER, I. *Über Aufbau und Wandlungen der Wahrnehmungswelt*. Vienna: Rohrer, 1951.

30. KÖHLER, W. An aspect of gestalt psychology. In Murchison, C. (Ed.), *Psychologies of 1925*. Worcester: Clark Univer. Press, 1926.

31. KÖHLER, W. *Dynamics in psychology*. New York: Liveright, 1940.

32. KÖHLER, W. *Gestalt psychology*. New York: Liveright, 1947.

33. KÖHLER, W., HELD, R., & O'CONNELL, D. N. An investigation of cortical currents. *Proc. Amer. phil. Soc.*, 1952, **96**, 291-330.

34. KÖHLER, W., & WALLACH, H. Figural after-effects. *Proc. Amer. phil. Soc.*, 1944, **88**, 269-357.

35. LASHLEY, K. S. Persistent problems in the evolution of mind. *Quart. Rev. Biol.*, 1949, **24**, 28-42.

36. LASHLEY, K. S., & RUSSELL, J. T. The mechanism of vision: XI. A preliminary test of innate organization. *J. genet. Psychol.*, 1934, **45**, 136-144.

37. LATTA, R. Notes on a case of successful operation for congenital cataract in an adult. *Brit. J. Psychol.*, 1904, **1**, 135-150.

38. LEEPER, R. A study of a neglected portion of the field of learning—the development of sensory organization. *J. genet. Psychol.*, 1935, **46**, 41-75.

39. LEVINE, R., CHEIN, I., & MURPHY, G. The relation of the intensity of a need to the amount of perceptual distortion: a preliminary report. *J. Psychol.*, 1942, **13**, 283-293.

40. LOCKE, N. M. Perception and intelligence: their phylogenetic relation. *Psychol. Rev.*, 1938, **45**, 335-345.

41. METZGER, W. Beobachtungen über phänomenale Identität. *Psychol. Forsch.*, 1934, **19**, 1-60.

42. METZGER, W. *Gesetze des Sehens*. (2nd ed.) Frankfurt am Main: Waldemar Kramer, 1953.

43. MICHOTTE, A. *La perception de la causalité*. (2nd ed.) Publications Universitaires de Louvain, 1954.

44. MILLER, M. Observation of initial visual experience in rats. *J. Psychol.*, 1948, **26**, 223-228.

45. MUNN, N. L. *Psychology*. (2nd ed.) Boston: Houghton Mifflin, 1951.

46. MURPHY, G. *Introduction to psychology*. New York: Harper, 1951.

47. NISSEN, H. W., CHOW, K. L., & SEMMES, J. Effects of restricted opportunity for tactful, kinesthetic, and manipulative experience on the behavior of a chimpanzee. *Amer. J. Psychol.*, 1951, **64**, 485-507.

48. OSGOOD, C. E. *Method and theory in experimental psychology*. New York: Oxford, 1953.

49. PRATT, C. C. The role of past experience in perception. *J. Psychol.*, 1950, **30**, 85-107.

50. PRATT, K. C. The neonate. In Carmichael, L. (Ed.), *Manual of child psychology*. (2nd ed.) New York: Wiley, 1954.

51. RIESEN, A. H. Arrested vision. *Sci. Amer.*, 1950, **183**, 16-19.

52. RIESEN, A. H. The development of visual perception in man and chimpanzee. *Science*, 1947, **106**, 107.

53. RIESEN, A. H. Plasticity of behavior: Psychological aspects. In Harlow, H. F., & Woolsey, C. N., *Symposium on biological and biochemical bases of behavior*. Madison, Wisconsin: Univer. of Wisconsin Press, in press.

54. RIESEN, A. H., KURKE, M. I., & MELLINGER, JEANNE C. Interocular transfer of habits learned monocularly in visually naive and visually experienced cats. *J. comp. physiol. Psychol.*, 1953, **46**, 166-172.

55. RIESEN, A. H., & MELLINGER, JEANNE C. Interocular transfer of habits in cats after alternating monocular visual experience. *J. comp. physiol. Psychol.*, 1956, **49**, 516-520.

56. ROCK, I., & FLECK, F. S. A re-examination of the effect of monetary reward and punishment in figure-ground perception. *J. exp. Psychol.*, 1950, **40**, 766-776.

57. ROCK, I., & KREMEN, I. A re-examination of Rubin's figural after-effect. *J. exp. Psychol.*, 1957, **53**, 23-30.

58. RUBIN, E. *Visuell wahrgenommene Figuren*. Copenhagen: 1921.

59. SCHAFER, R., & MURPHY, G. The role of autism in a figure-ground relationship. *J. exp. Psychol.*, 1943, 335-343.

60. SCHLODTMANN, W. Ein Beitrag zur Lehre von der optischen Localisation bei Blindgeborenen. *Arch. Ophthal.*, 1902, **54**, 256-267.

61. VON SENDEN, M. *Raum-und Gestaltauf-fassung bei operierten Blindgeborenen vor und nach der Operation.* Leipzig: Barth, 1932.
62. SHERMAN, M. C., SHERMAN, I. C., & FLORY, C. D. Infant behavior. *Comp. Psychol. Monogr.*, 1936, **12**, No. 4.
63. SIEGEL, A. I. Deprivation of visual form definition in the ring dove: I. Discriminatory learning. *J. comp. physiol. Psychol.*, 1953, **46**, 115-119.
64. SIEGEL, A. I. Deprivation of visual form definition in the ring dove: II. Perceptual-motor transfer. *J. comp. physiol. Psychol.*, 1953, **46**, 249-252.
65. SIEGEL, A. I. The effects of visual form definition upon the development and transfer of simple form discrimination in the ring dove. Ph.D. thesis, New York Univer., 1952.
66. SMITH, D. E., & HOCHBERG, J. E. The effect of "punishment" (electric shock) on figure-ground perception. *J. Psychol.*, 1954, **38**, 83-87.
67. SOLOMON, R. L., & HOWES, D. H. Word frequency, personal values, and visual duration thresholds. *Psychol. Rev.*, 1951, **58**, 256-270.
68. SPERRY, R. W. The eye and the brain. *Sci. Amer.*, 1956, **194**, 48-52.
69. TERNUS, J. Experimentelle Untersuchung über phänomene Identität. *Psychol. Forsch.*, 1926, **7**, 81-136.
70. THOMPSON, W. R., & MELZACK, R. Early environment. *Sci. Amer.*, 1956, **194**, 38-42.
71. TOCH, H. H., & ITTLESON, W. H. The role of past experience in apparent movement: a revaluation. *Brit. J. Psychol.*, 1956, **47**, 195-207.
72. WALLACH, H. Brightness constancy and the nature of achromatic colors. *J. exp. Psychol.*, 1948, **38**, 310-324.
73. WALLACH, H. On constancy of visual speed. *Psychol. Rev.*, 1939, **46**, 541-552.
74. WALLACH, H. The role of head movements and vestibular and visual cues in sound localization. *J. exp. Psychol.*, 1940, **27**, 339-368.
75. WALLACH, H. Some considerations concerning the relation between perception and cognition. *J. Pers.*, 1949, **18**, 6-13.
76. WALLACH, H., & ADAMS, PAULINE. Recognition and the localization of visual traces. *Amer. J. Psychol.*, 1954, **67**, 338-340.
77. WALLACH, H., & O'CONNELL, D. N. The kinetic depth effect. *J. exp. Psychol.*, 1953, **45**, 205-217.
78. WALLACH, H., O'CONNELL, D. N., & NEISER, U. The memory effect of visual perception of three-dimensional form. *J. exp. Psychol.*, 1953, **45**, 360-368.
79. WALLS, G. The problem of visual direction. *Amer. J. Optom. and Arch. Amer. Acad. Optom.*, 1951, Monogr. 117.
80. WERTHEIMER, MAX. Untersuchungen zur Lehre von der Gestalt, II. *Psychol. Forsch.*, 1923, **4**, 301-350.
81. WERTHEIMER, MICHAEL. Hebb and Senden on the role of learning in perception. *Amer. J. Psychol.*, 1951, **64**, 133-137.
82. WHEATSTONE, C. Contributions to the physiology of vision, I. *Phil. Trans., Pt. I*, 1838, 371-394.

Received December 3, 1956.

## FACTORS RELEVANT TO THE VALIDITY OF EXPERIMENTS IN SOCIAL SETTINGS<sup>1</sup>

DONALD T. CAMPBELL

*Northwestern University*

What do we seek to control in experimental designs? What extraneous variables which would otherwise confound our interpretation of the experiment do we wish to rule out? The present paper attempts a specification of the major categories of such extraneous variables and employs these categories in evaluating the validity of standard designs for experimentation in the social sciences.

Validity will be evaluated in terms of two major criteria. First, and as a basic minimum, is what can be called *internal validity*: did in fact the experimental stimulus make some significant difference in this specific instance? The second criterion is that of *external validity, representativeness, or generalizability*: to what populations, settings, and variables can this effect be generalized? Both criteria are obviously important although it turns out that they are to some extent incompatible, in that the controls required for internal validity often tend to jeopardize representativeness.

The extraneous variables affecting internal validity will be introduced in

<sup>1</sup> A dittoed version of this paper was privately distributed in 1953 under the title "Designs for Social Science Experiments." The author has had the opportunity to benefit from the careful reading and suggestions of L. S. Burwen, J. W. Cotton, C. P. Duncan, D. W. Fiske, C. I. Hovland, L. V. Jones, E. S. Marks, D. C. Pelz, and B. J. Underwood, among others, and wishes to express his appreciation. They have not had the opportunity of seeing the paper in its present form, and bear no responsibility for it. The author also wishes to thank S. A. Stouffer (33) and B. J. Underwood (36) for their public encouragement.

the process of analyzing three pre-experimental designs. In the subsequent evaluation of the applicability of three true experimental designs, factors leading to external invalidity will be introduced. The effects of these extraneous variables will be considered at two levels: as simple or main effects, they occur independently of or in addition to the effects of the experimental variable; as interactions, the effects appear in conjunction with the experimental variable. The main effects typically turn out to be relevant to internal validity, the interaction effects to external validity or representativeness.

The following designation for experimental designs will be used: *X* will represent the exposure of a group to the experimental variable or event, the effects of which are to be measured; *O* will refer to the process of observation or measurement, which can include watching what people do, listening, recording, interviewing, administering tests, counting lever depressions, etc. The *X*s and *O*s in a given row are applied to the same specific persons. The left to right dimension indicates temporal order. Parallel rows represent equivalent samples of persons unless otherwise specified. The designs will be numbered and named for cross-reference purposes.

### THREE PRE-EXPERIMENTAL DESIGNS AND THEIR CONFOUNDED EXTRANEous VARIABLES

*The One-Shot Case Study.* As Stouffer (32) has pointed out, much social science research still uses De-

sign 1, in which a single individual or group is studied in detail only once, and in which the observations are attributed to exposure to some prior situation.

*X O*      1. One-Shot Case Study

This design does not merit the title of experiment, and is introduced only to provide a reference point. The very minimum of useful scientific information involves at least one formal comparison and therefore at least two careful observations (2).

*The One-Group Pretest-Posttest Design.* This design does provide for one formal comparison of two observations, and is still widely used.

*O<sub>1</sub> X O<sub>2</sub>*    2. One-Group Pretest-Posttest Design

However, in it there are four or five categories of extraneous variables left uncontrolled which thus become rival explanations of any difference between  $O_1$  and  $O_2$ , confounded with the possible effect of  $X$ .

The first of these is the main effect of *history*. During the time span between  $O_1$  and  $O_2$  many events have occurred in addition to  $X$ , and the results might be attributed to these. Thus in Collier's (8) experiment, while his respondents<sup>2</sup> were reading Nazi propaganda materials, France fell, and the obtained attitude changes seemed more likely a result of this event than of the propaganda.<sup>3</sup> By history is meant the specific event series other than  $X$ , i.e., the extra-experimental uncontrolled stimuli. Relevant to this variable is the concept of experimental isolation, the employment of experimental settings

in which all extraneous stimuli are eliminated. The approximation of such control in much physical and biological research has permitted the satisfactory employment of Design 2. But in social psychology and the other social sciences, if history is confounded with  $X$  the results are generally uninterpretable.

The second class of variables confounded with  $X$  in Design 2 is here designated as *maturity*. This covers those effects which are systematic with the passage of time, and not, like history, a function of the specific events involved. Thus between  $O_1$  and  $O_2$  the respondents may have grown older, hungrier, tireder, etc., and these may have produced the difference between  $O_1$  and  $O_2$ , independently of  $X$ . While in the typical brief experiment in the psychology laboratory, maturation is unlikely to be a source of change, it has been a problem in research in child development and can be so in extended experiments in social psychology and education. In the form of "spontaneous remission" and the general processes of healing it becomes an important variable to control in medical research, psychotherapy, and social remediation.

There is a third source of variance that could explain the difference between  $O_1$  and  $O_2$  without a recourse to the effect of  $X$ . This is the effect of *testing* itself. It is often true that persons taking a test for the second time make scores systematically different from those taking the test for the first time. This is indeed the case for intelligence tests, where a second mean may be expected to run as much as five IQ points higher than the first one. This possibility makes important a distinction between *reactive* measures and *nonreactive* measures. A reactive measure is one

<sup>2</sup> In line with the central focus on social psychology and the social sciences, the term *respondent* is employed in place of the terms *subject*, *patient*, or *client*.

<sup>3</sup> Collier actually used a more adequate design than this, an approximation to Design 4.

which modifies the phenomenon under study, which changes the very thing that one is trying to measure. In general, any measurement procedure which makes the subject self-conscious or aware of the fact of the experiment can be suspected of being a reactive measurement. Whenever the measurement process is *not* a part of the normal environment it is probably reactive. Whenever measurement exercises the process under study, it is almost certainly reactive. Measurement of a person's height is relatively nonreactive. However, measurement of weight, introduced into an experimental design involving adult American women, would turn out to be reactive in that the process of measuring would stimulate weight reduction. A photograph of a crowd taken in secret from a second story window would be nonreactive, but a news photograph of the same scene might very well be reactive, in that the presence of the photographer would modify the behavior of people seeing themselves being photographed. In a factory, production records introduced for the purpose of an experiment would be reactive, but if such records were a regular part of the operating environment they would be nonreactive. An English anthropologist may be nonreactive as a participant-observer at an English wedding, but might be a highly reactive measuring instrument at a Dobu nuptials. Some measures are so extremely reactive that their use in a pretest-posttest design is not usually considered. In this class would be tests involving surprise, deception, rapid adaptation, or stress. Evidence is amply present that tests of learning and memory are highly reactive (35, 36). In the field of opinion and attitude research our well-developed interview and attitude test tech-

niques must be rated as reactive, as shown, for example, by Crespi's (9) evidence.

Even within the personality and attitude test domain, it may be found that tests differ in the degree to which they are reactive. For some purposes, tests involving voluntary self-description may turn out to be more reactive (especially at the interaction level to be discussed below) than are devices which focus the respondent upon describing the external world, or give him less latitude in describing himself (e.g., 5). It seems likely that, apart from considerations of validity, the Rorschach test is less reactive than the TAT or MMPI. Where the reactive nature of the testing process results from the focusing of attention on the experimental variable, it may be reduced by imbedding the relevant content in a comprehensive array of topics, as has regularly been done in Hovland's attitude change studies (14). It seems likely that with attention to the problem, observational and measurement techniques can be developed which are much less reactive than those now in use.

*Instrument decay* provides a fourth uncontrolled source of variance which could produce an  $O_1-O_2$  difference that might be mistaken for the effect of  $X$ . This variable can be exemplified by the fatiguing of a spring scales, or the condensation of water vapor in a cloud chamber. For psychology and the social sciences it becomes a particularly acute problem when human beings are used as a part of the measuring apparatus, as judges, observers, raters, coders, etc. Thus  $O_1$  and  $O_2$  may differ because the raters have become more experienced, more fatigued, have acquired a different adaptation level, or have learned about the purpose of the ex-

periment, etc. However infelicitously, this term will be used to typify those problems introduced when shifts in measurement conditions are confounded with the effect of  $X$ , including such crudities as having a different observer at  $O_1$  and  $O_2$ , or using a different interviewer or coder. Where the use of different interviewers, observers, or experimenters is unavoidable, but where they are used in large numbers, a sampling equivalence of interviewers is required, with the relevant  $N$  being the  $N$  of interviewers, not interviewees, except as refined through cluster sampling considerations (18).

A possible fifth extraneous factor deserves mention. This is statistical *regression*. When, in Design 2, the group under investigation has been selected for its extremity on  $O_1$ ,  $O_1-O_2$  shifts toward the mean will occur which are due to random imperfections of the measuring instrument or random instability within the population, as reflected in the test-retest reliability. In general, regression operates like maturation in that the effects increase systematically with the  $O_1-O_2$  time interval. McNemar (22) has demonstrated the profound mistakes in interpretation which failure to control this factor can introduce in remedial research.

*The Static Group Comparison.* The third pre-experimental design is the Static Group Comparison.

$X \quad O_1$   
----- 3. The Static Group Comparison  
             $O_2$

In this design, there is a comparison of a group which has experienced  $X$  with a group which has not, for the purpose of establishing the effect of  $X$ . In contrast with Design 6, there is in this design no means of certifying that the groups were equivalent

at some prior time. (The absence of sampling equivalence of groups is symbolized by the row of dashes.) This design has its most typical occurrence in the social sciences, and both its prevalence and its weakness have been well indicated by Stouffer (32). It will be recognized as one form of the correlational study. It is introduced here to complete the list of confounding factors. If the  $O$ s differ, this difference could have come about through biased *selection* or recruitment of the persons making up the groups; i.e., they might have differed anyway without the effect of  $X$ . Frequently, exposure to  $X$  (e.g., some mass communication) has been voluntary and the two groups have an inevitable systematic difference on the factors determining the choice involved, a difference which no amount of matching can remove.

A second variable confounded with the effect of  $X$  in this design can be called experimental *mortality*. Even if the groups were equivalent at some prior time,  $O_1$  and  $O_2$  may differ now not because individual members have changed, but because a biased subset of members have dropped out. This is a typical problem in making inferences from comparisons of the attitudes of college freshmen and college seniors, for example.

#### TRUE EXPERIMENTAL DESIGNS

*The Pretest-Posttest Control Group Design.* One or another of the above considerations led psychologists between 1900 and 1925 (2, 30) to expand Design 2 by the addition of a control group, resulting in Design 4.

$O_1 \ X \ O_2$  4. Pretest-Posttest Control Group  
             $O_3 \quad O_4$  Design

Because this design so neatly controls for the main effects of history, maturation, testing, instrument de-

cay, regression, selection, and mortality, these separate sources of variance are not usually made explicit. It seems well to state briefly the relationship of the design to each of these confounding factors, with particular attention to the application of the design in social settings.

If the differences between  $O_1$  and  $O_2$  were due to intervening historical events, then they should also show up in the  $O_3-O_4$  comparison. Note, however, several complications in achieving this control. If respondents are run in groups, and if there is only one experimental session and one control session, then there is no control over the unique internal histories of the groups. The  $O_1-O_3$  difference, even if not appearing in  $O_3-O_4$ , may be due to a chance distracting factor appearing in one or the other group. Such a design, while controlling for the shared history or event series, still confounds  $X$  with the unique session history. Second, the design implies a simultaneity of  $O_1$  with  $O_3$  and  $O_2$  with  $O_4$  which is usually impossible. If one were to try to achieve simultaneity by using two experimenters, one working with the experimental respondents, the other with the controls, this would confound experimenter differences with  $X$  (introducing one type of instrument decay). These considerations make it usually imperative that, for a true experiment, the experimental and control groups be tested and exposed individually or in small subgroups, and that sessions of both types be temporally and spatially intermixed.

As to the other factors: if maturation or testing contributed an  $O_1-O_2$  difference, this should appear equally in the  $O_3-O_4$  comparison, and these variables are thus controlled for their main effects. To make sure the design controls for instrument decay,

the same individual or small-session approximation to simultaneity needed for history is required. The occasional practice of running the experimental group and control group at different times is thus ruled out on this ground as well as that of history. Otherwise the observers may have become more experienced, more hurried, more careless, the maze more redolent with irrelevant cues, the lever-tension and friction diminished, etc. Only when groups are effectively simultaneous do these factors affect experimental and control groups alike. Where more than one experimenter or observer is used, counterbalancing experimenter, time, and group is recommended. The balanced Latin square is frequently useful for this purpose (4).

While regression is controlled in the design as a whole, frequently secondary analyses of effects are made for extreme pretest scorers in the experimental group. To provide a control for effects of regression, a parallel analysis of extremes should also be made for the control group.

Selection is of course handled by the sampling equivalence ensured through the randomization employed in assigning persons to groups, perhaps supplemented by, but not supplanted by, matching procedures. Where the experimental and control groups do not have this sort of equivalence, one has a compromise design rather than a true experiment. Furthermore, the  $O_1-O_3$  comparison provides a check on possible sampling differences.

The design also makes possible the examination of experimental mortality, which becomes a real problem for experiments extended over weeks or months. If the experimental and control groups do not differ in the number of lost cases nor in their

pretest scores, the experiment can be judged internally valid on this point, although mortality reduces the generalizability of effects to the original population from which the groups were selected.

For these reasons, the Pretest-Posttest Control Group Design has been the ideal in the social sciences for some thirty years. Recently, however, a serious and avoidable imperfection in it has been noted, perhaps first by Schanck and Goodman (29). Solomon (30) has expressed the point as an *interaction* effect of testing. In the terminology of analysis of variance, the effects of history, maturation, and testing, as described so far, are all *main* effects, manifesting themselves in mean differences independently of the presence of other variables. They are effects that could be added on to other effects, including the effect of the experimental variable. In contrast, interaction effects represent a joint effect, specific to the concomitance of two or more conditions, and may occur even when no main effects are present. Applied to the testing variable, the interaction effect might involve not a shift due solely or directly to the measurement process, but rather a sensitization of respondents to the experimental variable so that when *X* was preceded by *O* there would be a change, whereas both *X* and *O* would be without effect if occurring alone. In terms of the two types of validity, Design 4 is internally valid, offering an adequate basis for generalization to other sampling-equivalent *pretested* groups. But it has a serious and systematic weakness in representativeness in that it offers, strictly speaking, no basis for generalization to the *unpretested* population. And it is usually the *unpretested* larger universe

from which these samples were taken to which one wants to generalize.

A concrete example will help make this clearer. In the NORC study of a United Nations information campaign (31), two equivalent samples, of a thousand each, were drawn from the city's population. One of these samples was interviewed, following which the city of Cincinnati was subjected to an intensive publicity campaign using all the mass media of communication. This included special features in the newspapers and on the radio, bus cards, public lectures, etc. At the end of two months, the second sample of 1,000 was interviewed and the results compared with the first 1,000. There were no differences between the two groups except that the second group was somewhat more pessimistic about the likelihood of Russia's cooperating for world peace, a result which was attributed to history rather than to the publicity campaign. The second sample was no better informed about the United Nations nor had it noticed in particular the publicity campaign which had been going on. In connection with a program of research on panels and the reinterview problem, Paul Lazarsfeld and the Bureau of Applied Social Research arranged to have the initial sample reinterviewed at the same time as the second sample was interviewed, after the publicity campaign. This reinterviewed group showed significant attitude changes, a high degree of awareness of the campaign and important increases in information. The inference in this case is unmistakably that the initial interview had sensitized the persons interviewed to the topic of the United Nations, had raised in them a focus of awareness which made the subsequent publicity campaign effective for them

but for them only. This study and other studies clearly document the possibility of interaction effects which seriously limit our capacity to generalize from the pretested experimental group to the unpretested general population. Hovland (15) reports a general finding which is of the opposite nature but is, nonetheless, an indication of an interactive effect. In his Army studies the initial pretest served to reduce the effects of the experimental variable, presumably by creating a commitment to a given position. Crespi's (9) findings support this expectation. Solomon (30) reports two studies with school children in which a spelling pretest reduced the effects of a training period. But whatever the direction of the effect, this flaw in the Pretest-Posttest Control Group Design is serious for the purposes of the social scientist.

*The Solomon Four-Group Design.* It is Solomon's (30) suggestion to control this problem by adding to the traditional two-group experiment two unpretested groups as indicated in Design 5.

$O_1$	$X$	$O_2$	
$O_3$		$O_4$	5. Solomon Four-Group Design
$X$		$O_5$	

This Solomon Four-Group Design enables one both to control and measure both the main and interaction effects of testing and the main effects of a composite of maturation and history. It has become the new ideal design for social scientists. A word needs to be said about the appropriate statistical analysis. In Design 4, an efficient single test embodying the four measurements is achieved through computing for each individual a pretest-posttest difference score which is then used for compar-

ing by  $t$  test the experimental and control groups. Extension of this mode of analysis to the Solomon Four-Group Design introduces an inelegant awkwardness to the otherwise elegant procedure. It involves assuming as a pretest score for the unpretested groups the mean value of the pretest from the first two groups. This restricts the effective degrees of freedom, violates assumptions of independence, and leaves one without a legitimate base for testing the significance of main effects and interaction. An alternative analysis is available which avoids the assumed pretest scores. Note that the four posttests form a simple two-by-two analysis of variance design:

	No $X$	$X$
Pretested	$O_1$	$O_2$
Unpretested	$O_3$	$O_4$

The column means represent the main effect of  $X$ , the row means the main effect of pretesting, and the interaction term the interaction of pretesting and  $X$ . (By use of a  $t$  test the combined main effects of maturation and history can be tested through comparing  $O_1$  with  $O_3$  and  $O_2$ .)

*The Posttest-Only Control Group Design.* While the statistical procedures of analysis of variance introduced by Fisher (10) are dominant in psychology and the other social sciences today, it is little noted in our discussions of experimental arrangements that Fisher's typical agricultural experiment involves no pretest: equivalent plots of ground receive different experimental treatments and the subsequent yields are measured.<sup>4</sup> Applied to a social experiment

<sup>4</sup> This is not to imply that the pretest is totally absent from Fisher's designs. He suggests the use of previous year's yields, etc., in covariance analysis. He notes, however, "with

as in testing the influence of a motion picture upon attitudes, two randomly assigned audiences would be selected, one exposed to the movie, and the attitudes of each measured subsequently for the first time.

*A X O<sub>1</sub>* 6. Posttest-Only Control Group  
*A O<sub>2</sub>* Design

In this design the symbol *A* had been added, to indicate that at a specific time prior to *X* the groups were made equivalent by a random sampling assignment. *A* is the point of selection, the point of allocation of individuals to groups. It is the existence of this process that distinguishes Design 6 from Design 3, the Static Group Comparison. Design 6 is not a static cross-sectional comparison, but instead truly involves control and observation extended in time. The sampling procedures employed assure us that at time *A* the groups were equal, even if not measured. *A* provides a point of prior equality just as does the pretest. A point *A* is, of course, involved in all true experiments, and should perhaps be indicated in Designs 4 and 5. It is essential that *A* be regarded as a specific point in time, for groups change as a function of time since *A*, through experimental mortality. Thus in a public opinion survey situation employing probability sampling from lists of residents, the longer the time since *A*, the more the sample underrepresents the transient segments of society, the newer dwelling units, etc. When experimental groups are being drawn from a self-selected extreme population, such as

annual agricultural crops, knowledge of yields of the experimental area in a previous year under uniform treatment has not been found sufficiently to increase the precision to warrant the adoption of such uniformity trials as a preliminary to projected experiments" (10, p. 176).

applicants for psychotherapy, time since *A* introduces maturation (spontaneous remission) and regression factors. In Design 6 these effects would be confounded with the effect of *X* if the *As* as well as the *Os* were not contemporaneous for experimental and control groups.

Like Design 4, this design controls for the effects of maturation and history through the practical simultaneity of both the *As* and the *Os*. In superiority over Design 4, no main or interaction effects of pretesting are involved. It is this feature that recommends it in particular. While it controls for the main and interaction effects of pretesting as well as does Design 5, the Solomon Four-Group Design, it does not measure these effects, nor the main effect of history-maturation. It can be noted that Design 6 can be considered as the two unpretested "control" groups from the Solomon Design, and that Solomon's two traditional pretested groups have in this sense the sole purpose of measuring the effects of pretesting and history-maturation, a purpose irrelevant to the main aim of studying the effect of *X* (25). However, under normal conditions of not quite perfect sampling control, the four-group design provides in addition greater assurance against mistakenly attributing to *X* effects which are not due it, inasmuch as the effect of *X* is documented in three different fashions (*O<sub>1</sub>* vs. *O<sub>2</sub>*, *O<sub>2</sub>* vs. *O<sub>4</sub>*, and *O<sub>6</sub>* vs. *O<sub>6</sub>*). But, short of the four-group design, Design 6 is often to be preferred to Design 4, and is a fully valid experimental design.

Design 6 has indeed been used in the social sciences, perhaps first of all in the classic experiment by Gosnell, *Getting Out the Vote* (11). Schanck and Goodman (29), Hovland (15) and others (1, 12, 23, 24, 27) have also

employed it. But, in spite of its manifest advantages of simplicity and control, it is far from being a popular design in social research and indeed is usually relegated to an inferior position in discussions of experimental designs if mentioned at all (e.g., 15, 16, 32). Why is this the case?

In the first place, it is often confused with Design 3. Even where Ss have been carefully assigned to experimental and control groups, one is apt to have an uneasiness about the design because one "doesn't know what the subjects were like before." This objection must be rejected, as our standard tests of significance are designed precisely to evaluate the likelihood of differences occurring by chance in such sample selection. It is true, however, that this design is particularly vulnerable to selection bias and where random assignment is not possible it remains suspect. Where naturally aggregated units, such as classes, are employed intact, these should be used in large numbers and assigned at random to the experimental and control conditions; cluster sampling statistics (18) should be used to determine the error term. If but one or two intact classrooms are available for each experimental treatment, Design 4 should certainly be used in preference.

A second objection to Design 6, in comparison with Design 4, is that it often has less precision. The difference scores of Design 4 are less variable than the posttest scores of Design 6 if there is a pretest-posttest correlation above .50 (15, p. 323), and hence for test-retest correlations above that level a smaller mean difference would be statistically significant for Design 4 than for Design 6, for a constant number of cases. This advantage to Design 4 may often be

more than dissipated by the costs and loss in experimental efficiency resulting from the requirement of two testing sessions, over and above the considerations of representativeness.

Design 4 has a particular advantage over Design 6 if experimental mortality is high. In Design 4, one can examine the pretest scores of lost cases in both experimental and control groups and check on their comparability. In the absence of this in Design 6, the possibility is opened for a mean difference resulting from differential mortality rather than from individual change, if there is a substantial loss of cases.

A final objection comes from those who wish to study the relationship of pretest attitudes to kind and amount of change. This is a valid objection, and where this is the interest, Design 4 or 5 should be used, with parallel analysis of experimental and control groups. Another common type of individual difference study involves classifying persons in terms of amount of change and finding associated characteristics such as sex, age, education, etc. While unavailable in this form in Design 6, essentially the same correlational information can be obtained by subdividing both experimental and control groups in terms of the associated characteristics, and examining the experimental-control difference for such subtypes.

For Design 6, the Posttest-Only Control Group Design, there is a class of social settings in which it is optimally feasible, settings which should be more used than they now are. Whenever the social contact represented by  $X$  is made to single individuals or to small groups, and where the response to that stimulus can be identified in terms of individuals or type of  $X$ , Design 6 can be applied. Direct mail and door-to-

door contacts represent such settings. The alternation of several appeals from door-to-door in a fund-raising campaign can be organized as a true experiment without increasing the cost of the solicitation. Experimental variation of persuasive materials in a direct-mail sales campaign can provide a better experimental laboratory for the study of mass communication and persuasion than is available in any university. The well-established, if little-used, split-run technique in comparing alternative magazine ads is a true experiment of this type, usually limited to coupon returns rather than sales because of the problem of identifying response with stimulus type (20). The split-ballot technique (7) long used in public opinion polls to compare different question wordings or question sequences provides an excellent example which can obviously be extended to other topics (e.g., 12). By and large these laboratories have not yet been used to study social science theories, but they are directly relevant to hypotheses about social persuasion.

*Multiple X designs.* In presenting the above designs,  $X$  has been opposed to No- $X$ , as is traditional in discussions of experimental design in psychology. But while this may be a legitimate description of the stimulus-isolated physical science laboratory, it can only be a convenient shorthand in the social sciences, for any No- $X$  period will not be empty of potentially change-inducing stimuli. The experience of the control group might better be categorized as another type of  $X$ , a control experience, an  $X_C$  instead of No- $X$ . It is also typical of advance in science that we are soon no longer interested in the qualitative fact of effect or no-effect, but want to specify degree of effect for varying degrees of  $X$ . These con-

siderations lead into designs in which multiple groups are used, each with a different  $X_1$ ,  $X_2$ ,  $X_3$ ,  $X_n$ , or in multiple factorial design, as  $X_{1a}$ ,  $X_{1b}$ ,  $X_{2a}$ ,  $X_{2b}$ , etc. Applied to Designs 4 and 6, this introduces one additional group for each additional  $X$ . Applied to 5, The Solomon Four-Group Design, two additional groups (one pretested, one not, both receiving  $X_n$ ) would be added for each variant on  $X$ .

In many experiments,  $X_1$ ,  $X_2$ ,  $X_3$ , and  $X_n$  are all given to the same group, differing groups receiving the  $X$ s in different orders. Where the problem under study centers around the effects of order or combination, such counterbalanced multiple  $X$  arrangements are, of course, essential. Studies of transfer in learning are a case in point (34). But where one wishes to generalize to the effect of each  $X$  as occurring in isolation, such designs are not recommended because of the sizable interactions among  $X$ s, as repeatedly demonstrated in learning studies under such labels as proactive inhibition and learning sets. The use of counterbalanced sets of multiple  $X$ s to achieve experimental equation, where natural groups not randomly assembled have to be used, will be discussed in a subsequent paper on compromise designs.

*Testing for effects extended in time.* The researches of Hovland and his associates (14, 15) have indicated repeatedly that the longer range effects of persuasive  $X$ s may be qualitatively as well as quantitatively different from immediate effects. These results emphasize the importance of designing experiments to measure the effect of  $X$  at extended periods of time. As the misleading early research on reminiscence and on the consolidation of the memory trace indicate (36), repeated measure-

ment of the same persons cannot be trusted to do this if a reactive measurement process is involved. Thus, for Designs 4 and 6, two separate groups must be added for each posttest period. The additional control group cannot be omitted, or the effects of intervening history, maturation, instrument decay, regression, and mortality are confounded with the delayed effects of  $X$ . To follow fully the logic of Design 5, four additional groups are required for each posttest period.

*True experiments in which  $O$  is not under  $E$ 's control.* It seems well to call the attention of the social scientist to one class of true experiments which are possible without the full experimental control over both the "when" and "to whom" of both  $X$  and  $O$ . As far as this analysis has been able to go, no such true experiments are possible without the ability to control  $X$ , to withhold it from carefully randomly selected respondents while presenting it to others. But control over  $O$  does not seem so indispensable. Consider the following design.

$A \quad X \quad O_1$   
 $A \quad O_2$  6. Posttest Only Design, where  $O$   
( $O$ ) cannot be withheld from any  
( $O$ ) respondent  
( $O$ )

The parenthetical  $O$ s are inserted to indicate that the studied groups, experimental and control, have been selected from a larger universe all of which will get  $O$  anyway. An election provides such an  $O$ , and using "whether voted" rather than "how voted," this was Gosnell's design (11). Equated groups were selected at time  $A$ , and the experimental group subjected to persuasive materials designed to get out the vote. Using precincts rather than persons as the basic sampling unit, similar

studies can be made on the content of the voting (6). Essential to this design is the ability to create specified randomly equated groups, the ability to expose one of these groups to  $X$  while withholding it (or providing  $X_2$ ) from the other group, and the ability to identify the performance of each individual or unit in the subsequent  $O$ . Since such measures are natural parts of the environment to which one wishes to generalize, they are not reactive, and Design 4, the Pretest-Posttest Control Group Design, is feasible if  $O$  has a predictable periodicity to it. With the precinct as a unit, this was the design of Hartmann's classic study of emotional vs. rational appeals in a public election (13). Note that 5, the Solomon Four-Group Design, is not available, as it requires the ability to withhold  $O$  experimentally, as well as  $X$ .

#### FURTHER PROBLEMS OF REPRESENTATIVENESS

The interaction effect of testing, affecting the external validity or representativeness of the experiment, was treated extensively in the previous section, inasmuch as it was involved in the comparison of alternative designs. The present section deals with the effects upon representativeness of other variables which, while equally serious, can apply to any of the experimental designs.

*The interaction effects of selection.* Even though the true experiments control selection and mortality for internal validity purposes, these factors have, in addition, an important bearing on representativeness. There is always the possibility that the obtained effects are specific to the experimental population and do not hold true for the populations to which one wants to generalize. Defining the universe of reference in advance and

selecting the experimental and control groups from this "at random" would guarantee representativeness if it were ever achieved in practice. But inevitably not all those so designated are actually eligible for selection by any contact procedure. Our best survey sampling techniques, for example, can designate for potential contact only those available through residences. And, even of those so designated, up to 19 per cent are not contactable for an interview in their own homes even with five callbacks (37). It seems legitimate to assume that the more effort and time required of the respondent, the larger the loss through nonavailability and noncooperation. If one were to try to assemble experimental groups away from their own homes it seems reasonable to estimate a 50 per cent selection loss. If, still trying to extrapolate to the general public, one further limits oneself to docile preassembled groups, as in schools, military units, studio audiences, etc., the proportion of the universe systematically excluded through the sampling process must approach 90 per cent or more. Many of the selection factors involved are indubitably highly systematic. Under these extreme selection losses, it seems reasonable to suspect that the experimental groups might show reactions not characteristic of the general population. This point seems worth stressing lest we unwarrantedly assume that the selection loss for experiments is comparable to that found for survey interviews in the home at the respondent's convenience. Furthermore, it seems plausible that the greater the cooperation required, the more the respondent has to deviate from the normal course of daily events, the greater will be the possibility of nonrepresentative reactions. By and large,

Design 6 might be expected to require less cooperation than Design 4 or 5, especially in the natural individual contact setting. The interactive effects of experimental mortality are of similar nature. Note that, on these grounds, the longer the experiment is extended in time the more respondents are lost and the less representative are the groups of the original universe.

*Reactive arrangements.* In any of the experimental designs, the respondents can become aware that they are participating in an experiment, and this awareness can have an interactive effect, in creating reactions to *X* which would not occur had *X* been encountered without this "I'm a guinea pig" attitude. Lazarsfeld (19), Kerr (17), and Rosenthal and Frank (28), all have provided valuable discussions of this problem. Such effects limit generalizations to respondents having this awareness, and preclude generalization to the population encountering *X* with non-experimental attitudes. The direction of the effect may be one of negativism, such as an unwillingness to admit to any persuasion or change. This would be comparable to the absence of any immediate effect from discredited communicators, as found by Hovland (14). The result is probably more often a cooperative responsiveness, in which the respondent accepts the experimenter's expectations and provides pseudoconfirmation. Particularly is this positive response likely when the respondents are self-selected seekers after the cure that *X* may offer. The Hawthorne studies (21), illustrate such sympathetic changes due to awareness of experimentation rather than to the specific nature of *X*. In some settings it is possible to disguise the experimental purpose by providing plausi-

ble façades in which  $X$  appears as an incidental part of the background (e.g., 26, 27, 29). We can also make more extensive use of experiments taking place in the intact social situation, in which the respondent is not aware of the experimentation at all.

The discussion of the effects of selection on representativeness has argued against employing intact natural preassembled groups, but the issue of conspicuousness of arrangements argues for such use. The machinery of breaking up natural groups such as departments, squads, and classrooms into randomly assigned experimental and control groups is a source of reaction which can often be avoided by the use of preassembled groups, particularly in educational settings. Of course, as has been indicated, this requires the use of large numbers of such groups under both experimental and control conditions.

The problem of reactive arrangements is distributed over all features of the experiment which can draw the attention of the respondent to the fact of experimentation and its purposes. The conspicuous or reactive pretest is particularly vulnerable, inasmuch as it signals the topics and purposes of the experimenter. For communications of obviously persuasive aim, the experimenter's topical intent is signaled by the  $X$  itself, if the communication does not seem a part of the natural environment. Even for the posttest-only groups, the occurrence of the posttest may create a reactive effect. The respondent may say to himself, "Aha, now I see why we got that movie." This consideration justifies the practice of disguising the connection between  $O$  and  $X$  even for Design 6, as through having different experimental per-

sonnel involved, using different façades, separating the settings and times, and embedding the  $X$ -relevant content of  $O$  among a disguising variety of other topics.<sup>8</sup>

*Generalizing to other Xs.* After the internal validity of an experiment has been established, after a dependable effect of  $X$  upon  $O$  has been found, the next step is to establish the limits and relevant dimensions of generalization not only in terms of populations and settings but also in terms of categories and aspects of  $X$ . The actual  $X$  in any one experiment is a specific combination of stimuli, all confounded for interpretative purposes, and only some relevant to the experimenter's intent and theory. Subsequent experimentation should be designed to purify  $X$ , to discover that aspect of the original conglomerate  $X$  which is responsible for the effect. As Brunswik (3) has emphasized, the representative sampling of  $X$ s is as relevant a problem in linking experiment to theory as is the sampling of respondents. To define a category of  $X$ s along some dimension, and then to sample  $X$ s for experimental purposes from the full range of stimuli meeting the specification while other aspects of each specific stimulus complex are varied, serves to untie or unconfound the defined dimension from specific others, lending assurance of theoretical relevance.

In a sense, the placebo problem can be understood in these terms. The

<sup>8</sup> For purposes of completeness, the interaction of  $X$  with history and maturation should be mentioned. Both affect the generalizability of results. The interaction effect of history represents the possible specificity of results to a given historical moment, a possibility which increases as problems are more societal, less biological. The interaction of maturation and  $X$  would be represented in the specificity of effects to certain maturational levels, fatigue states, etc.

experiment without the placebo has clearly demonstrated that some aspect of the total  $X$  stimulus complex has had an effect; the placebo experiment serves to break up the complex  $X$  into the suggestive connotation of pill-taking and the specific pharmacological properties of the drug—separating two aspects of the  $X$  previously confounded. Subsequent studies may discover with similar logic which chemical fragment of the complex natural herb is most essential. Still more clearly, the sham operation illustrates the process of  $X$  purification, ruling out general effects of surgical shock so that the specific effects of loss of glandular or neural tissue may be isolated. As these parallels suggest, once recurrent unwanted aspects of complex  $X$ s have been discovered for a given field, control groups especially designed to eliminate these effects can be regularly employed.

*Generalizing to other Os.* In parallel form, the scientist in practice uses a complex measurement procedure which needs to be refined in subsequent experimentation. Again, this is best done by employing multiple  $O$ s all having in common the theoretically relevant attribute but varying widely in their irrelevant specificities. For  $O$ s this process can be introduced into the initial experiment by employing multiple measures. A major practical reason for not doing so is that it is so frequently a frustrating experience, lending hesitancy, indecision, and a feeling of failure to studies that would have been interpreted with confidence had but a single response measure been employed.

*Transition experiments.* The two previous paragraphs have argued against the *exact* replication of experimental apparatus and measurement procedures on the grounds that this

continues the confounding of theory-relevant aspects of  $X$  and  $O$  with specific artifacts of unknown influence. On the other hand, the confusion in our literature generated by the heterogeneity of results from studies all on what is nominally the "same" problem but varying in implementation, is leading some to call for exact replication of initial procedures in subsequent research on a topic. Certainly no science can emerge without dependably repeatable experiments. A suggested resolution is the *transition experiment*, in which the need for varying the theory-independent aspects of  $X$  and  $O$  is met in the form of a multiple  $X$ , multiple  $O$  design, one segment of which is an "exact" replication of the original experiment, exact at least in those major features which are normally reported in experimental writings.

*Internal vs. external validity.* If one is in a situation where either internal validity or representativeness must be sacrificed, which should it be? The answer is clear. Internal validity is the prior and indispensable consideration. The optimal design is, of course, one having both internal and external validity. Insofar as such settings are available, they should be exploited, without embarrassment from the apparent opportunistic warping of the content of studies by the availability of laboratory techniques. In this sense, a science is as opportunistic as a bacteria culture and grows only where growth is possible. One basic necessity for such growth is the machinery for selecting among alternative hypotheses, no matter how limited those hypotheses may have to be.

#### SUMMARY

In analyzing the extraneous variables which experimental designs for

social settings seek to control, seven categories have been distinguished: history, maturation, testing, instrument decay, regression, selection, and mortality. In general, the simple or main effects of these variables jeopardize the internal validity of the experiment and are adequately controlled in standard experimental designs. The interactive effects of these variables and of experimental arrangements affect the external valid-

ity or generalizability of experimental results. Standard experimental designs vary in their susceptibility to these interactive effects. Stress is also placed upon the differences among measuring instruments and arrangements in the extent to which they create unwanted interactions. The value for social science purposes of the Posttest-Only Control Group Design is emphasized.

## REFERENCES

- ANNIS, A. D., & MEIER, N. C. The induction of opinion through suggestion by means of planted content. *J. soc. Psychol.*, 1934, **5**, 65-81.
- BORING, E. G. The nature and history of experimental control. *Amer. J. Psychol.*, 1954, **67**, 573-589.
- BRUNSWIK, E. *Perception and the representative design of psychological experiments*. Berkeley: Univer. of California Press, 1956.
- BUGELSKI, B. R. A note on Grant's discussion of the Latin square principle in the design and analysis of psychological experiments. *Psychol. Bull.*, 1949, **46**, 49-50.
- CAMPBELL, D. T. The indirect assessment of social attitudes. *Psychol. Bull.*, 1950, **47**, 15-38.
- CAMPBELL, D. T. On the possibility of experimenting with the "Bandwagon" effect. *Int. J. Opin. Attitude Res.*, 1951, **5**, 251-260.
- CANTRIL, H. *Gauging public opinion*. Princeton: Princeton Univer. Press, 1944.
- COLLIER, R. M. The effect of propaganda upon attitude following a critical examination of the propaganda itself. *J. soc. Psychol.*, 1944, **20**, 3-17.
- CRESPI, L. P. The interview effect in polling. *Publ. Opin. Quart.*, 1948, **12**, 99-111.
- FISHER, R. A. *The design of experiments*. Edinburgh: Oliver & Boyd, 1935.
- GOSNELL, H. F. *Getting out the vote: an experiment in the stimulation of voting*. Chicago: Univer. of Chicago Press, 1927.
- GREENBERG, A. Matched samples. *J. Marketing*, 1953-54, **18**, 241-245.
- HARTMANN, G. W. A field experiment on the comparative effectiveness of "emotional" and "rational" political leaflets
- in determining election results. *J. abnorm. soc. Psychol.*, 1936, **31**, 99-114.
- HOVLAND, C. E., JANIS, I. L., & KELLEY, H. H. *Communication and persuasion*. New Haven: Yale Univer. Press, 1953.
- HOVLAND, C. I., LUMSDAINE, A. A., & SHEFFIELD, F. D. *Experiments on mass communication*. Princeton: Princeton Univer. Press, 1949.
- JAHODA, M., DEUTSCH, M., & COOK, S. W. *Research methods in social relations*. New York: Dryden Press, 1951.
- KERR, W. A. Experiments on the effect of music on factory production. *Appl. Psychol. Monogr.*, 1945, No. 5.
- KISH, L. Selection of the sample. In L. Festinger and D. Katz (Eds.), *Research methods in the behavioral sciences*. New York: Dryden Press, 1953, 175-239.
- LAZARSFELD, P. F. Training guide on the controlled experiment in social research. Dittoed. Columbia Univer., Bureau of Applied Social Research, 1948.
- LUCAS, D. B., & BRITT, S. H. *Advertising psychology and research*. New York: McGraw-Hill, 1950.
- MAYO, E. *The human problems of an industrial civilization*. New York: Macmillan, 1933.
- MCNEMAR, Q. A critical examination of the University of Iowa studies of environmental influences upon the IQ. *Psychol. Bull.*, 1940, **37**, 63-92.
- MENEFE, S. C. An experimental study of strike propaganda. *Soc. Forces*, 1938, **16**, 574-582.
- PARRISH, J. A., & CAMPBELL, D. T. Measuring propaganda effects with direct and indirect attitude tests. *J. abnorm. soc. Psychol.*, 1953, **48**, 3-9.
- PAYNE, S. L. The ideal model for con-

trolled experiments. *Publ. Opin. Quart.*, 1951, 15, 557-562.

26. POSTMAN, L., & BRUNER, J. S. Perception under stress. *Psychol. Rev.*, 1948, 55, 314-322.

27. RANKIN, R. E., & CAMPBELL, D. T. Galvanic skin response to Negro and white experimenters. *J. abnorm. soc. Psychol.*, 1955, 51, 30-33.

28. ROSENTHAL, D., & FRANK, J. O. Psychotherapy and the placebo effect. *Psychol. Bull.*, 1956, 53, 294-302.

29. SCHANCK, R. L., & GOODMAN, C. Reactions to propaganda on both sides of a controversial issue. *Publ. Opin. Quart.*, 1939, 3, 107-112.

30. SOLOMON, R. W. An extension of control group design. *Psychol. Bull.*, 1949, 46, 137-150.

31. STAR, S. A., & HUGHES, H. M. Report on an educational campaign: the Cincinnati plan for the United Nations. *Amer. J. Sociol.*, 1949-50, 55, 389.

32. STOUFFER, S. A. Some observations on study design. *Amer. J. Sociol.*, 1949-50, 55, 355-361.

33. STOUFFER, S. A. Measurement in sociology. *Amer. sociol. Rev.*, 1953, 18, 591-597.

34. UNDERWOOD, B. J. *Experimental psychology*. Appleton-Century-Crofts, 1949.

35. UNDERWOOD, B. J. Interference and forgetting. *Psychol. Rev.*, 1957, 64, 49-60.

36. UNDERWOOD, B. J. *Psychological research*. New York: Appleton-Century-Crofts, 1957.

37. WILLIAMS, R. Probability sampling in the field: a case history. *Publ. Opin. Quart.*, 1950, 14, 316-330.

Received October 7, 1956

## TRANSFER DESIGNS AND FORMULAS<sup>1</sup>

BENNET B. MURDOCK, JR.

*University of Vermont*

Transfer of training has been defined as ". . . the effect of a *preceding* activity upon the learning of a given task" (58, p. 520). The preceding activity will be referred to here as Task 1. The task to be learned, that to which the transfer occurs, will be referred to as Task 2. This paper discusses the experimental designs and formulas that are appropriate in two types of transfer studies, verbal transfer (VT) and predifferentiation (PD). Studies of VT are those in which the material to be learned in both Task 1 and Task 2 is verbal, usually nonsense syllables. Studies of PD are those in which Ss in Task 1 become familiar with the stimuli in one of several possible ways (for instance, by studying them individually, by noting similarities or differences, or by learning distinctive labels for them); then in Task 2 the effects of this familiarization are determined by having Ss learn simple discriminative motor responses to these same stimuli.<sup>2</sup> Typical experiments in VT have been performed by Bruce (8) and Underwood (74); typical experiments in PD have been performed by Gagné and Baker (24) and Rossman and Goss (63).

Studies of VT and PD are considered together because they have much the same problems of methodology. Not only are they both transfer studies but also the type of learning involved in Task 2, where the test for transfer occurs, is very simi-

lar. While on Task 2 in VT Ss may be calling out nonsense syllables and in PD they may be pressing buttons, still in both cases Ss are learning which of *X* responses is correct for each of *Y* stimuli (*X* and *Y* usually being the same). That is, Ss are learning simple discriminative responses to a number of different stimuli.

On the other hand, studies of motor transfer usually investigate the acquisition of a rather complex motor response, and this presents a somewhat different set of methodological problems from those encountered in studying the learning of simple discriminative responses. Of course, it is sometimes hard to differentiate between PD studies and motor transfer studies, especially when both types of studies may use tasks which the authors refer to as "perceptual-motor." As used here the distinction between PD and motor transfer is this: if the second task response is a simple one which has been learned prior to the experiment (pressing a button, throwing a switch) this would be PD; if it is the response itself rather than the association of a response to a stimulus that is learned (mirror drawing, pursuit rotor) this would be motor transfer.

This paper, then, deals with experimental designs and formulas that are appropriate to studies of VT and PD but not to studies of motor transfer.<sup>3</sup>

<sup>1</sup> The preparation of this article was in part supported by a research grant, NSF G2590, from the National Science Foundation.

<sup>2</sup> A few PD studies (3, 55, 62) have investigated the effects of the familiarization process on the recognition of the stimuli rather than on learning, and these studies will also be included.

<sup>3</sup> Also, for the most part this paper will exclude studies of stimulus generalization (which deal with the evocation, not the acquisition, of a specific response in the second task), studies of retroactive and proactive inhibition (which are more concerned with retention than with learning), and transposition experiments (which use a transfer paradigm to infer from Task 2 what was learned in Task 1).

Although there have been a few PD studies carried out on the animal level this report will be restricted to studies which have dealt with human learning.

#### TRANSFER DESIGNS

Most discussions of transfer designs are closely modeled after those originally suggested by Woodworth (75, 76). Woodworth listed five different designs, and these differed primarily in two dimensions, the use of equated tasks or equated groups, and the use of before-after tests or successive practice. Most experiments in VT and PD use equated groups; the use of equated tasks is usually limited to ensuring that various first tasks are equivalent to each other or that various second tasks are equivalent to each other rather than dealing with equivalent first and second tasks. While the use of a before test (on the second task) is usually essential in motor transfer, it is seldom if ever used in VT or PD. Rather the method of successive practice is almost universal; Ss start from scratch on Task 2 and continue until some criterion is reached.

It would seem, then, that in studies of VT and PD the most common design is one utilizing equated groups and successive practice. This is the one listed by Woodworth as Plan 4 (75, p. 180), where the experimental group learns Task 1 and then Task 2 while the control group learns Task 2 only. However, a survey of recent VT and PD experiments showed that there are other designs in use, even though some of them have not been explicitly recognized. The following section, then, will list five different designs that are currently in use and will evaluate them primarily from the point of view of their validity.

*Design I.* (29, 38, 62.) All Ss learn

Task 1 and then Task 2. Either the two tasks are equated or, with counterbalancing, half the Ss learn Task 2 and then Task 1. This design has been used, for instance, to determine if there is more transfer going from tactile to visual stimuli or from visual to tactile stimuli (29).

*Design II.* (2, 7, 10, 11, 12, 35, 53, 63, 69.) All Ss learn the same Task 1 and the same Task 2 but differ in the time interval or interpolated activity between the two tasks. This design has been used to determine the amount of transfer occurring over varying periods of time (10) or to determine the effects of warm-up on Task 2 learning (69).

*Design III.* (1, 3, 5, 6, 8, 11, 13, 14, 15, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 30, 32, 34, 39, 40, 41, 44, 46, 47, 48, 49, 51, 56, 60, 62, 63, 65, 67, 68, 70, 71, 73, 74, 78.) The Ss in the experimental group (E group) learn Task 1 and then Task 2; Ss in the control group (C group) engage in a different preliminary activity, Task 1', before learning Task 2.<sup>4</sup> This is Woodworth's Plan 4, the one that is often designated as the standard design. Design III has, for instance, been used to determine if serial learning (Task 2) occurs more rapidly with prior familiarization (Task 1) than without any prior familiarization (Task 1') (40).

<sup>4</sup> As Osgood (58) has clearly pointed out, if we compare an E group having Task 1 and then Task 2 with a C group having Task 2 only we are comparing the effect of a specific preceding activity with that of a nonspecific preceding activity. In other words, prior to learning Task 2, Ss in the C group are not literally doing nothing; they are resting, judging cartoons, naming colors, or on their way to the experiment. Therefore, this nonspecific preceding activity is referred to here as Task 1'. Of course, Task 1' does not have to be so nonspecific; sometimes its nature will be clearly indicated, as when it consists of having Ss learn material similar to Task 1 but irrelevant to Task 2.

This design could be (though as pointed out seldom is) modified by introducing a foretest on Task 2 to ensure equality of groups or to match them. So modified, Design III would be identical to the standard RI design. The foretest would correspond to the original learning, Tasks 1 and 1' to the interpolated activity, and Task 2 to the test for retention. However, in RI studies the original learning is usually carried to a higher degree than would be desirable for a foretest in a transfer experiment.

*Design IV.* (2, 6, 7, 9, 16, 17, 19, 31, 34, 36, 37, 40, 42, 46, 47, 49, 54, 55, 57, 59, 60, 61, 63, 64, 65, 66.) The Ss in the E group learn Task 1 and then Task 2; Ss in the C group learn Task 1 and then a similar, though not identical, second task, Task 2'. Design IV has been used to show that, following the learning of A—B (Task 1), it is easier to learn A—D (Task 2) than it is to learn A—X (Task 2') given a B—C—D chain of pre-established associations (64).

To illustrate the difference between Design III and Design IV: We wish to determine if learning labels for stimuli facilitates the acquisition of simple motor responses to these same stimuli. For stimuli we select colors of various shades of blue and colors of various shades of green. In Design III Ss in the E group learn both labels and motor responses to the blue colors while Ss in the C group learn labels for the green colors but motor responses to the blue. In Design IV Ss in the E group learn both labels and motor responses to the blue colors, but Ss in the C group learn labels for the blue colors and motor responses to the green.

*Design V.* (4, 8, 33, 38, 43, 52, 74, 77, 78.) The Ss in one group learn Task 1 and then Task 2; Ss in the second group learn both a different first task, Task 1', and a different

second task, Task 2'. As usually used Tasks 1 and 1' are similar and Tasks 2 and 2' are also similar; however, the intertask relationship between Tasks 1 and 2 is different from that between Tasks 1' and 2'. In one experiment in which this design was used the interlist response similarity was varied from none to synonymity (52).

It is, of course, axiomatic that all transfer depends upon intertask relationships. In Design V it is quite apparent that the intertask relationship differs for the various groups. However, it also differs in Designs III and IV, in the first case because the first task varies and in the second case because the second task varies. Actually, the difference among the five designs is in where the experimental variation is introduced. In Design I it is in the sequence in which the two tasks are learned. In Design II it is in the period intervening between the first and second tasks. In Design III the variation is in the first task, in Design IV in the second task, and in Design V it is in both.

There is, of course, no reason to restrict the designs to two groups; this has been done in an attempt to simplify the descriptions. In studying transfer over varying periods of time (Design II) many different time intervals may be used. Design III is a common one to study the effects of different degrees of Task 1 learning; many experiments use four groups with no, low, medium, and high first-list learning. Design IV is often used to study S—R position in paired-associate VT; following an A—B list there may be an A—C, C—B, C—D, and an A—B rearranged list. With Design V a number of different degrees of interlist similarity may be used.

Many experiments study the effects of two or more independent variables

and use different designs for the different variables. Therefore, many experiments were listed under more than one design. A fairly common combination is to use Design III to study the effects of degree of first-task learning and at the same time Design IV to study the effects of S-R position. An experiment by Rossman and Goss (63) actually used three different designs; Design II for shock or no shock, Design III for number of trials of Task 1 learning, and Design IV for verbalization or no verbalization on the second task.

*Evaluation of the designs.* In his evaluation of the methods used in the study of human learning and retention Melton (50) listed three criteria to use in evaluating any particular methodology: validity, reliability, and conformity. Validity refers to the extent to which the results obtained can unequivocally be ascribed to the experimental variable or variables under investigation. Reliability refers to the consistency and conformity to the standardization of methods. Since validity is probably the *sine qua non* of methodology, and since validity is one of the recurring problems in the design of transfer experiments, this is the chief criterion which will be used in evaluating the five transfer designs listed above. Since Designs III, IV, and V are probably the most common and the most important designs in VT and PD, they will be considered first.

The basic problem in Design III is ensuring that Task 1 and Task 1' differ in one and only one way; i.e., differ only in the experimental variable under investigation. And two reasons why this is often a real problem are warm-up and learning-how-to-learn. A number of experiments have shown that the learning of a given task is effected, and sometimes markedly, both by the immediately preceding

activity and by the number (or extent) of similar tasks previously learned (45). If then the E group learns Task 1 and then Task 2 while the C group "rests" and then learns Task 2 any resulting differences in the learning of Task 2 may be due to the specific effects of the particular Task 1, the warm-up from Task 1, or the general facilitation from learning a similar, though not identical, prior task.

This, then, is perhaps the clearest case of an invalid design. From such an experiment we could, of course, draw conclusions about the over-all direction and degree of transfer. However, we would not know the relative weights to assign the three factors as determinants of the transfer. Of course, if the net transfer were negative we could be sure that there was interference from the specific Task 1, but we wouldn't know how much greater it would have been had warm-up and learning-how-to-learn been controlled.

One paradoxical aspect of this situation is that Design III is almost universally used in studying the one variable where there is the most reason to assume that warm-up and learning-how-to-learn would have a definite effect; that is, in studying the effect of the degree of Task 1 learning on transfer. In the typical experiment of this type the various groups differ in the amount of practice on Task 1, and the almost universal result is that transfer increases as practice on Task 1 increases. Yet it is reasonable to assume that both warm-up and learning-how-to-learn also increase with practice on Task 1 and it may be this rather than the specific facilitation resulting from better mastery of Task 1 that accounts for the positive transfer. To control for these general practice effects it is necessary for all groups to have the same over-all

amount of practice on a prior task but to vary the amount of practice on a relevant first task. Only two experiments have been found which have controlled for warm-up and learning-how-to-learn; one of them (56) found the usual facilitation while the other (15) did not. Perhaps it would be wiser to suspend judgment about the effects of the degree of familiarization on transfer until more adequately controlled experiments have been performed.

There are other situations where these general practice effects may be confounded with the experimental variable. For instance, if Task 1 involves learning responses to stimuli while Task 1' involves studying them, for instance, to note similarities and differences there may be more warm-up or a better developed learning set resulting from Task 1. Probably the ideal solution to this problem is to use for Tasks 1 and 1' two tasks which are basically identical but differ in that one (Task 1) is relevant to Task 2 while the other (Task 1') is not.

Design IV does not have this problem because the same first task is used for both E and C groups. The chief problem with this design is ensuring that Tasks 2 and 2' are equivalent; i.e., would be equally difficult to learn in the absence of Task 1. If this assumption cannot be made then it is impossible to draw any conclusions whatsoever about transfer; differences between the E and C group may simply reflect the fact that one second task is more difficult than the other. Without appropriate controls then this would also be an invalid design.

It is, of course, possible to determine if Task 2 and Task 2' are of equal difficulty by having a separate control group learn both second tasks without Task 1. Or as another alter-

native it is possible to control for possible inequalities between the two second tasks. One way that this could be done which is particularly suitable for studies of VT would be to get one master second list and then subdivide it into Task 2 and Task 2' by randomly assigning the S-R pairs to one of the two tasks. This is, of course, directly comparable to the generally accepted procedure of randomly assigning Ss as a means of obtaining equal groups. A second method of control is to use counterbalancing. With counterbalancing for one group of Ss Task 2 would be the experimental task and Task 2' the control task. This would be reversed for the second group. Any systematic differences should balance out and thus not affect transfer.

Counterbalancing, however, immediately introduces an additional factor. There must be two first tasks, Task 1 as a relevant first task when Task 2 is the experimental task and Task 1' as a relevant first task when Task 2' is the experimental task. We then need four groups, Group A to learn Task 1 and Task 2, B to learn 1 and 2', C to learn 1' and 2', and D to learn 1' and 2. Thus, B serves as a baseline to measure the transfer for A, and D the baseline for C. The over-all measure of transfer would then be  $(A - B) + (C - D)$ .

Of course, counterbalancing can also be used with Design III. In the logic of this Design D serves as the baseline for A, and B as the baseline for C. The over-all measure of transfer here would be  $(A - D) + (C - B)$ . Since  $(A - B) + (C - D) = (A - D) + (C - B)$  it can be argued that, with counterbalancing, Designs III and IV are identical. However, this is not necessarily so; with Design IV each S can serve as his own control (and this is one of the big advantages of Design IV over Design III). With

each *S* serving as his own control only two groups of *Ss* are necessary, one group to learn Task 1 and then both Tasks 2 and 2' while a second group learns Task 1' and again both Tasks 2 and 2'. Actually in some cases Tasks 2 and 2' can be scrambled together and presented as one task though scored as two (9, 42, 47, 54, 57, 64). The comparable procedure with Design III would be for one group to learn both Tasks 1 and 1' then Task 2 while a second group learns Tasks 1 and 1' and then Task 2'. This of course doesn't make sense; there is no group which learns a second task without previously having learned a relevant first task.

At first glance it would appear that Design V had the problems of both Designs III and IV. Since one group will learn Task 1 while a second group learns Task 1' there is the problem of ensuring that the first tasks differ in only one way. And since one group will learn Task 2 and a second group Task 2' there is the additional problem of ensuring that the two second tasks are of equal difficulty. However some type of counterbalancing is almost always used with this design (72) and the counterbalancing may handle these problems adequately.

If the order in which *Ss* experience treatments is counterbalanced this will control for learning-how-to-learn (though if the various first tasks require different amounts of practice this does not control for warm-up). If the specific tasks are counterbalanced among treatments (i.e., if a given task is Task 1 half the time and Task 1' the other half) possible inequalities among the various first tasks will be controlled. Even if the tasks are not counterbalanced it is at least possible to determine from the results if the various first tasks are comparable and, if they are not, the conclusions can be modified accordingly.

If the specific tasks are counterbalanced among treatments on the second task this would be the same type of control discussed in connection with Design IV. If they are not; if one task is always Task 2, another one Task 2', and so on, then there is more of a problem. Within the framework of the transfer experiment itself it is not possible to determine if the various second tasks are of equal difficulty, because task difficulty is confounded with transfer effects. Here there should either be some compelling reason to believe on a priori grounds that the various second tasks are equivalent or a separate control group should be run to determine this empirically.

In Design I if all *Ss* learn Task 1 and then Task 2 practice effects and the specific facilitation from Task 1 are confounded. To have a valid design with counterbalancing it is probably necessary to assume that practice effects from Task 1 to Task 2 are equal to the practice effects from Task 2 to Task 1. Unless this assumption seems reasonable it would probably be safer, though more complicated, to use either Design III or IV.

One possible use of Design I with counterbalancing would be to test the basic assumption for Design IV; that is, that Tasks 2 and 2' are equivalent. This is one case where it probably would be reasonable to assume that the practice effects in both directions were comparable. The other way of testing the equality of the two tasks would be, of course, to have two separate groups, one to learn Task 2 and the other to learn Task 2'.

Of the five designs, Design II is probably the one with the fewest difficulties. Both groups learn the same first and second tasks, so there is no problem about ensuring equality. In Design II the chief concern is that the intervening activities differ in

only one way, but this need not present any unusual problems of control.

*Selection of appropriate design.* In selecting an appropriate design for a particular transfer experiment the validity of the design is, of course, of particular importance. However, the particular problem to be studied is also an important determinant. If, for instance, one wishes to determine if the transfer from A to B differs from the transfer from B to A, Design I is the logical choice. Or in studying warm-up itself or the nature of the intervening activity Design II would be used.

Since transfer is presumably a function both of the task from which transfer occurs and the task to which transfer occurs either of these is a fit subject for investigation. On logical grounds the first would utilize Design III and the second Design IV. However it is probably not necessary to adhere strictly to this principle. Both designs (as well as the others) are different ways of studying the same basic problem of transfer, the effect of different intertask relationships. Therefore if a different design seemed more suitable on other groups (especially validity) it should probably be used.

Finally, when each *S* is tested under a number of (or all) different conditions Design V is probably necessary. Obviously the same *S* cannot learn the same task more than once, and Design V is the only one that provides a number of separate first and second tasks.

#### TRANSFER FORMULAS

This section deals with the problem of determining the amount of transfer obtained in a given experiment. If Design I is used without counterbalancing, the basic comparison is be-

tween the learning of Task 1 and the learning of Task 2. In all other designs the basic comparison is between the learning of Task 2 by the E group and the learning of Task 2 or Task 2' by the C group. If the measure of learning used is such that the larger the numerical value the better the learning (as would be true with number of correct responses) the amount of transfer would be represented by (E-C). If the measure of learning is such that the larger the value the poorer the learning (number of errors or number of trials to reach criterion) the amount of transfer would be given by (C-E). In this way the sign of the difference would indicate whether the transfer was positive or negative.

Each of these measures, (E-C) and (C-E), yields a clear-cut measure of the amount of transfer obtained in a particular experiment. However, as has been pointed out (28), since the values are in raw score units it is impossible with this measure to compare the results of experiments which have used different measures of learning. What is needed, then, is a measure of transfer which is independent of the raw score units. For studies of VT and PD probably the best way of doing this is to express the difference between E and C as a percentage.

In their article on the measurement of transfer of training Gagné, Foster, and Crowley (28) suggest two different ways of obtaining value indicating the percentage of transfer. The first way to do this is to compare the difference between the E and C groups with the performance of the C group itself. If the measure of learning is one such as number of correct responses the formula would be:

$$\text{Percentage of transfer} = \frac{E - C}{C} \times 100. \quad [1a]$$

If the measure of learning is one such as number of errors or trials the formula would be:

$$\text{Percentage of transfer} = \frac{C - E}{C} \times 100. \quad [1b]$$

The second way is to compare the difference between the E and C groups to the maximum amount of improvement possible. The maximum improvement possible is determined by the difference between the total possible score on Task 2 (here indicated by T) and the actual performance of the C group on Task 2. Thus if the measure of learning is one such as the number of correct responses the formula would be:

$$\text{Percentage of transfer} = \frac{E - C}{T - C} \times 100. \quad [2a]$$

If the measure of learning is one such as number of errors or trials the formula would be:

$$\text{Percentage of transfer} = \frac{C - E}{C - T} \times 100. \quad [2b]$$

That these two types of formulas really do differ can be seen by the following hypothetical example: In Experiment A,  $T = 20$ ,  $E = 16$ , and  $C = 12$ ; in Experiment B,  $T = 80$ ,  $E = 25$ , and  $C = 15$ . For which Experiment, A or B, is there the greater transfer? By Formula [1a] the answer would be B, 67% to 33%; by Formula [2a] the answer would be A, 50% to 15%. These two formulas differ not only as to which experiment showed the greater transfer but also as to

\* Gagné *et al.* label the first type of formula "per cent improvement" and the second type "per cent transfer." Actually, the former reflects improvement (of the E group over the C group) relative to the C group while the latter reflects improvement relative to the maximum improvement possible. Here, however, both will be referred to as percentage of transfer.

the absolute amount of transfer in each. Clearly then it makes a difference whether the difference between the E group and the C group is compared to the performance of the C group or to the total amount of improvement that is possible.

Gagné *et al.* make a strong case for the second type of formula. They feel that for both theoretical and practice purposes it is more desirable to determine how close the obtained transfer comes to the maximum amount of transfer that is possible than to determine how great the transfer is relative to the level of zero transfer (which is given by the performance of the C group on Task 2). It is certainly true that at times it would be very desirable to determine how great the transfer is relative to the maximum possible. However, there are at least four reasons why the use of the second type of formula may not be completely satisfactory.

1. Determination of T may be difficult or impossible. Is T to be considered perfect performance (i.e., no errors, no trials to learn, or 100% correct responses)? If so, this would seem quite unrealistic; even a group which had perfected Task 2 prior to being tested on it would usually not exhibit perfect performance on the test for transfer. If not, then T would presumably have to be determined empirically—as Gagné and Foster (26, 27) actually do in the two cases in which they use this type of formula. Then, however, total possible score becomes the best score obtained, and what started out as a theoretical limit becomes an empirical limit. This difficulty of getting an appropriate value for T is probably the single greatest weakness in the use of the second type of formula.

2. If T is empirically determined

there might be times when its validity could be questioned. What guarantee could there be that any given group actually performed at its best? Or, if all groups in the experiment were poorer than the C group one might be forced to the rather strange conclusion that T was given by the performance of the group showing the least negative transfer.

3. Although it is generally considered that transfer effects decrease as Task 2 learning progresses, the second type of formula may show transfer effects to increase toward the later stages of learning. This would occur, for instance, with Formula [2a] if the differences between the E and C groups were approximately the same at the beginning, middle, and end of learning and if the difference between T and C decreased as learning continued (which is probably to be expected). Under the same conditions Formula [1a] would show transfer effects to decrease, as the denominator would be increasing.

4. As Gagné *et al.* point out (28, pp. 104-105) these formulas are unsatisfactory for negative transfer. The lower limit of both is minus infinity, not -100%. Also, -100% transfer is not comparable to +100% transfer; i.e., the latter indicates the best performance possible but the former does not indicate the worst performance possible. Of course, this criticism also applies to the first type of formula as well.

These then are four difficulties which may arise in using the second type of formula. With the exception of the last one these problems would not apply to the first type of formula as there is no reference to total possible score. As for the first type of formula there are two main criticisms which Gagné *et al.* seem to make of it.

1. "Their [Formulas 1a and 1b]

outstanding limitation is the fact that percent improvement is a measure which is dependent upon the raw score units, and does not permit a comparison with the percent improvement obtained with other tasks" (p. 106). However, as has been pointed out, the first type of formula expresses transfer as a percentage, and two percentages can always be compared. If one experiment used number of correct responses as the measure of learning and found 50% transfer while a second experiment used number of trials to reach criterion as the measure of learning and found only 25% transfer one could still say that there was more transfer in the first experiment. Of course, one could not be sure whether the greater transfer found in the first experiment was a function of different measures of learning or greater facilitation from the first task, but still the direction of the difference is unambiguous.

Later on Gagné *et al.* claim that, "The [Formulas 2a and 2b] yield a measure of transfer which is independent of variations in the rate of learning of different tasks employed in transfer experiments, and thus permit comparisons to be made between studies" (p. 112). Perhaps what Gagné *et al.* mean when they say that the outstanding limitation of the first type of formula is that it is "dependent upon the raw score units" is that it is dependent upon variations in the rate of learning of the second task. In this connection they state (p. 101) that, with the first type of formula, comparisons among the percentage of transfer obtained at different stages of Task 2 learning are unjustifiable unless the negatively accelerated portion of the Task 2 learning curve is taken into consideration.

It is certainly true that the shape of the Task 2 learning curve (i.e.,

"variations in the rate of learning") will affect the amount of transfer obtained if the first type of formula is used. However, it would seem that the shape of the Task 2 learning curve would also affect transfer if the second type of formula were used. Thus, with a negatively accelerated curve in Task 2 the C group will probably be far from maximum performance early in learning but close to the limit late in learning, and this will directly determine the values that the denominator will take in the second type of formula.

Probably both types of formulas are in part dependent upon the shape of the Task 2 learning curve. However, it can be argued that this is a strength, not a weakness, of both types of formulas. As has been pointed out, transfer effects are a function not only of the task from which transfer occurs but also of the task to which transfer occurs, and it would seem desirable that a transfer formula be sensitive to both sets of variables.

2. The second criticism which Gagné *et al.* make of the first type of formula seems to be this: percentage-wise a large difference between E and C groups may be misleading if the performance of both is quite poor relative to T. This is certainly true; to take an example even more extreme than theirs; if  $T = 105$ ,  $E = 10$ , and  $C = 5$ , to call this 100% transfer by Formula [1a] neglects the fact that even the E group is very poor relative to what could be achieved. On the other hand, to call this 5% transfer by Formula [2a] seems equally misleading—after all, the E group is twice as good as the C group, poor though both may be.

It has been suggested\* that what

Gagné *et al.* were attempting to find was a measure of transfer independent of the raw score units in the sense that the standard deviation is independent of raw score units: with a normal distribution not only does one sigma above and below the mean, for instance, include about two-thirds of the group irrespective of the units of the distribution but also a standard score of +1.00, for instance, is just as far above the mean as a standard score of -1.00 is below the mean. With a transfer formula that was independent of raw score units in this sense a Task 1 that resulted in 50% positive transfer in one situation would be just as effective as a Task 1 that resulted in 50% positive transfer in a different situation. Also, a Task 1 producing 50% positive transfer would facilitate Task 2 learning just as much as a Task 1 that produced 50% negative transfer would interfere with it. To meet these requirements a transfer formula would have to be such that positive and negative transfer were symmetrical, and to be symmetrical the absolute value of the upper and lower limits must be identical (preferably, of course, 100%). It is on this last point, identical upper and lower limits, that both types of transfer formulas are unsatisfactory.

There is one way to modify the first type of formula so that positive and negative transfer would be symmetrical and the upper and lower limits would be 100%. This is to make the denominator include the performance of the E group as well as the performance of the C group. If the measure of learning were number of correct responses the formula would be:

[3a]

$$\text{Percentage of transfer} = \frac{E - C}{E + C} \times 100.$$

\* The author would like to thank Miss Harriet Foster for this and several other very helpful suggestions.

If the measure of learning were number of errors or trials the formula would be:

$$\text{Percentage of transfer} = \frac{C - E}{E + C} \times 100. \quad [3b]$$

To compare the three types of transfer formulas Table 1 lists a number of hypothetical results and the

TABLE 1  
PERCENTAGE OF TRANSFER AS DETERMINED  
BY EACH OF THE THREE FORMULAS

Number of Correct Responses			Formulas		
E	C	T	[1a]	[2a]	[3a]
20	0	20	+ ∞	+ 100%	+ 100%
15	5	20	+ 200%	+ 67%	+ 50%
10	5	20	+ 100%	+ 33%	+ 33%
5	5	20	0%	0%	0%
5	10	20	- 50%	- 50%	- 33%
5	15	20	- 67%	- 200%	- 50%
0	20	20	- 100%	- ∞	- 100%

percentage of transfer given by Formulas [1a], [2a], and [3a]. In this particular example Formula [1a] is clearly more suitable for negative transfer than for positive transfer, as in the latter case the transfer can go to plus infinity. The opposite holds true for Formula [2a] where the negative transfer can go to minus infinity. Only Formula [3a] is symmetrical and has an upper and lower

limit of 100%. However it should be noted that in general Formula [3a] gives smaller values than either of the other two formulas. In using this formula on some of his own data it has been the author's experience that the third type of formula usually does give rather small values for the percentage of transfer obtained.

Here, then, are three different transfer formulas, each with its own advantages and disadvantages. It does seem that the third type may be preferable to the first type. Whether or not it is preferable to the second type probably depends primarily on how important it is to try to relate the obtained transfer to the maximum transfer possible.

In conclusion, even though none of the formulas is perfect it is better to use some formula than none at all. Of some fifty-eight experimental studies published since the Gagné article appeared in 1948 only five (6, 26, 27, 37, 54) used any formula whatsoever, and of these five two (26, 27) were part of the original Gagné series. If we are to make real progress in establishing functional relationships in the area of transfer it is absolutely essential, as Gagné et al. (28) clearly point out, to have some means of representing the amount of transfer so as to compare different studies. That is why it is necessary to develop and use transfer formulas.

#### REFERENCES

- ADAMS, J. A. Psychomotor response acquisition and transfer as a function of control-indicator relationships. *J. exp. Psychol.*, 1954, 48, 10-14.
- ANDREWS, T. G., SHAPIRO, S., & COFER, C. N. Transfer and generalization of the inhibitory potential developed in rote serial learning. *Amer. J. Psychol.*, 1954, 67, 453-463.
- ARNOUlt, M. D. Transfer of pre-differentiation training in simple and multiple shape discrimination. *J. exp. Psychol.*, 1953, 45, 401-409.
- ATWATER, S. K. Proactive inhibition and associative facilitation as affected by degree of prior learning. *J. exp. Psychol.*, 1953, 46, 400-404.
- BAKER, KATHERINE E., & WYLIE, RUTH C. Transfer of verbal training to a motor task. *J. exp. Psychol.*, 1950, 40, 632-638.
- BATTIG, W. F. Transfer from verbal pre-

training to motor performance as a function of motor task complexity. *J. exp. Psychol.*, 1956, **51**, 371-378.

7. BIRGE, JANE S. Verbal responses in transfer. Unpublished doctor's dissertation, Yale Univer., 1941.
8. BRUCE, R. W. Conditions of transfer of training. *J. exp. Psychol.*, 1933, **16**, 343-361.
9. BUGELSKI, B. R., & SCHARLOECK, D. P. An experimental demonstration of unconscious mediated association. *J. exp. Psychol.*, 1952, **44**, 334-338.
10. BUNCH, MARION E. The amount of transfer in relational learning as a function of time. *J. comp. Psychol.*, 1936, **22**, 325-337.
11. BUNCH, MARION E. Cumulative transfer of training under different temporal conditions. *J. comp. Psychol.*, 1944, **37**, 265-272.
12. BUNCH, MARION E., & McCRAVEN, V. G. The temporal course of transfer in the learning of memory material. *J. comp. Psychol.*, 1938, **25**, 481-496.
13. BUNCH, MARION E., & WINSTON, M. M. The relationship between the character of the transfer and retroactive inhibition. *Amer. J. Psychol.*, 1936, **48**, 598-608.
14. CANTOR, G. N. Effects of three types of pretraining on discrimination learning in preschool children. *J. exp. Psychol.*, 1955, **49**, 339-342.
15. CANTOR, JOAN H. Amount of pretraining as a factor in stimulus predifferentiation and performance set. *J. exp. Psychol.*, 1955, **50**, 180-184.
16. CASTENADA, A. Effects of stress on complex learning and performance. *J. exp. Psychol.*, 1956, **52**, 9-12.
17. CASTENADA, A., & PALERMO, D. S. Psychomotor performance as a function of amount of training and stress. *J. exp. Psychol.*, 1955, **50**, 175-179.
18. DIETZ, DORIS. The facilitating effect of words on discrimination and generalization. *J. exp. Psychol.*, 1955, **50**, 255-260.
19. DUNCAN, C. P. Transfer in motor learning as a function of degree of first-task learning and inter-task similarity. *J. exp. Psychol.*, 1953, **45**, 1-11.
20. DYSINGER, D. W. An investigation of stimulus pre-differentiation in a choice discrimination problem. Unpublished doctor's dissertation, State Univer. of Iowa, 1951.
21. ECKSTRAND, G. A., & WICKENS, D. D. Transfer of perceptual set. *J. exp. Psychol.*, 1954, **47**, 274-278.
22. FARBER, I. E., & MURFIN, F. L. Performance set as a factor in transfer of training. Paper read at Midwest. Psychol. Ass., Chicago, April, 1951.
23. FOSTER, HARRIET. Stimulus predifferentiation in transfer of training. Unpublished doctor's dissertation, Univer. of Michigan, 1953.
24. GAGNÉ, R. M., & BAKER, KATHERINE E. Stimulus predifferentiation as a factor in transfer of training. *J. exp. Psychol.*, 1950, **40**, 439-451.
25. GAGNÉ, R. M., BAKER, KATHERINE E., & FOSTER, HARRIET. Transfer of discrimination training to a motor task. *J. exp. Psychol.*, 1950, **40**, 314-328.
26. GAGNÉ, R. M., & FOSTER, HARRIET. Transfer of training from practice on components in a motor skill. *J. exp. Psychol.*, 1949, **39**, 47-68.
27. GAGNÉ, R. M., & FOSTER, HARRIET. Transfer to a motor skill from practice on a pictured representation. *J. exp. Psychol.*, 1949, **39**, 342-354.
28. GAGNÉ, R. M., FOSTER, HARRIET, & CROWLEY, MIRIAM E. The measurement of transfer of training. *Psychol. Bull.*, 1948, **45**, 97-130.
29. GAYDOS, H. F. Intersensory transfer in the discrimination of form. *Amer. J. Psychol.*, 1956, **69**, 107-110.
30. GERJUOY, IRMA R. Discrimination learning as a function of the similarity of the stimulus names. Unpublished doctor's dissertation, State Univer. of Iowa, 1953.
31. GIBSON, ELEANOR J. Retroactive inhibition as a function of degree of generalization between tasks. *J. exp. Psychol.*, 1941, **28**, 93-115.
32. GOSS, A. E. Transfer as a function of type and amount of preliminary experience with task stimuli. *J. exp. Psychol.*, 1953, **46**, 419-428.
33. GREGG, L. W. The effect of stimulus complexity on discriminative responses. *J. exp. Psychol.*, 1954, **48**, 289-297.
34. HAKE, H. H., & ERIKSEN, C. W. Effect of number of permissible response categories on learning of a constant number of visual stimuli. *J. exp. Psychol.*, 1955, **50**, 161-167.
35. HAMILTON, C. E. The relationship between length of interval separating two learning tasks and the performance on the second. *J. exp. Psychol.*, 1950, **40**, 613-621.
36. HAMILTON, R. JANE. Retroactive facilitation as a function of degree of generalization between tasks. *J. exp. Psychol.*, 1943, **32**, 363-376.

37. HARcum, E. R. Verbal transfer of overlearned forward and backward associations. *Amer. J. Psychol.*, 1953, **66**, 622-625.

38. HERON, W. T. Warming-up effect in learning nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 219-228.

39. HOLTOn, RUTH B., & GOSS, A. E. Transfer to a discriminative motor task as a function of amount and type of preliminary verbalization. *J. gen. Psychol.*, 1956, **55**, 117-126.

40. HOVLAND, C. I., & KURTZ, K. H. Experimental studies in rote-learning theory: X. Pre-learning syllable familiarization and the length-difficulty relationship. *J. exp. Psychol.*, 1952, **44**, 31-39.

41. JEFFREY, W. E. The effects of verbal and nonverbal responses in mediating an instrumental act. *J. exp. Psychol.*, 1953, **45**, 327-333.

42. KURTZ, K. H. Discrimination of complex stimuli: The relationship of training and test stimuli in transfer of discrimination. *J. exp. Psychol.*, 1955, **50**, 283-292.

43. L'ABATE, L. Transfer and manifest anxiety in paired-associate learning. *Psychol. Rep.*, 1956, **2**, 119-126.

44. McALLISTER, DOROTHY E. The effects of various kinds of relevant verbal pre-training on subsequent motor performance. *J. exp. Psychol.*, 1953, **46**, 329-336.

45. McGEOCH, J. A., & IRION, A. L. *The psychology of human learning*. (2nd. ed.) New York: Longmans, Green, 1952.

46. MALTZMAN, I., & BROOKS, L. O. A failure to find second-order semantic generalization. *J. exp. Psychol.*, 1956, **51**, 413-417.

47. MANDLER, G. Transfer of training as a function of degree of response overlearning. *J. exp. Psychol.*, 1954, **49**, 411-417.

48. MANDLER, G. The warm-up effect: Some further evidence on temporal and task factors. *J. gen. Psychol.*, 1956, **55**, 3-8.

49. MANDLER, G., & HEINEMANN, SHIRLEY H. Effect of overlearning of a verbal response on transfer of training. *J. exp. Psychol.*, 1956, **52**, 39-46.

50. MELTON, A. W. The methodology of experimental studies of human learning and retention. I. The functions of a methodology and the available criteria for evaluating different experimental methods. *Psychol. Bull.*, 1936, **33**, 305-394.

51. MELTON, A. W., & IRWIN, J. M. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *Amer. J. Psychol.*, 1940, **53**, 173-203.

52. MORGAN, R. L., & UNDERWOOD, B. J. Proactive inhibition as a function of response similarity. *J. exp. Psychol.*, 1950, **40**, 592-603.

53. MURDOCK, B. B., JR. The effects of failure and retroactive inhibition on mediated generalization. *J. exp. Psychol.*, 1952, **44**, 156-164.

54. MURDOCK, B. B., JR. "Backward" learning in paired associates. *J. exp. Psychol.*, 1956, **51**, 213-215.

55. NEISSER, U. An experimental distinction between perceptual process and verbal response. *J. exp. Psychol.*, 1954, **47**, 399-402.

56. NOBLE, C. E. The effect of familiarization upon serial verbal learning. *J. exp. Psychol.*, 1955, **49**, 333-338.

57. OSGOOD, C. E. Meaningful similarity and interference in learning. *J. exp. Psychol.*, 1946, **36**, 277-301.

58. OSGOOD, C. E. *Method and theory in experimental psychology*. New York: Oxford Univer. Press, 1953.

59. PORTER, L. W., & DUNCAN, C. P. Negative transfer in verbal learning. *J. exp. Psychol.*, 1953, **46**, 61-64.

60. PRICE, HELEN G., & LEWIS, D. Increased pronouncing behavior as a factor in serial learning. *J. exp. Psychol.*, 1954, **47**, 95-100.

61. ROBINSON, IRENE P. The effects of differential degrees of similarity of stimulus-response relations on transfer of verbal learning. *Amer. Psychologist*, 1948, **3**, 250. (Abstract)

62. ROBINSON, J. S. The effect of learning verbal labels for stimuli on their later discrimination. *J. exp. Psychol.*, 1955, **49**, 112-114.

63. ROSSMAN, IRMA L., & GOSS, A. E. The acquired distinctiveness of cues: The role of discriminative verbal responses in facilitating the acquisition of discriminative motor responses. *J. exp. Psychol.*, 1951, **42**, 173-182.

64. RUSSELL, W. A., & STORMS, L. H. Implicit verbal chaining in paired-associates learning. *J. exp. Psychol.*, 1955, **49**, 287-293.

65. SHEFFIELD, F. D. The role of meaningfulness of stimuli and responses in verbal learning. Unpublished doctor's dissertation, Yale Univer., 1946.

66. SMITH, M. H., JR. Instructional sets and habit interference. *J. exp. Psychol.*, 1952, **44**, 267-272.

67. SMITH, S. L., & GOSS, A. E. The role of

the acquired distinctiveness of cues in the acquisition of a motor skill in children. *J. genet. Psychol.*, 1955, **87**, 11-24.

68. SPIKER, C. C. Stimulus pretraining and subsequent performance in the delayed reaction experiment. *J. exp. Psychol.*, 1956, **52**, 107-111.

69. THUNE, L. E. The effect of different types of preliminary activities on subsequent learning of paired-associate material. *J. exp. Psychol.*, 1950, **40**, 423-438.

70. THUNE, L. E. Warm-up effect as a function of level of practice in verbal learning. *J. exp. Psychol.*, 1951, **42**, 250-256.

71. UNDERWOOD, B. J. Associative inhibition in the learning of successive paired-associate lists. *J. exp. Psychol.*, 1944, **34**, 127-135.

72. UNDERWOOD, B. J. *Experimental psychology*. New York: Appleton-Century-Crofts, 1949.

73. UNDERWOOD, B. J. Proactive inhibition as a function of time and degree of prior learning. *J. exp. Psychol.*, 1949, **39**, 24-34.

74. UNDERWOOD, B. J. Associative transfer in verbal learning as a function of response similarity and degree of first-list learning. *J. exp. Psychol.*, 1951, **42**, 44-53.

75. WOODWORTH, R. S. *Experimental psychology*. New York: Holt, 1938.

76. WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology*. (Rev. Ed.) New York: Holt, 1954.

77. YOUNG, R. K. Retroactive and proactive effects under varying conditions of response similarity. *J. exp. Psychol.*, 1955, **50**, 113-119.

78. YOUNG, R. K., & UNDERWOOD, B. J. Transfer in verbal materials with dissimilar stimuli and response similarity varied. *J. exp. Psychol.*, 1954, **47**, 153-159.

Received October 22, 1956.

## EXPERIMENTAL STUDIES ON FIGURAL AFTEREFFECTS IN JAPAN

MORIJI SAGARA

*University of Tokyo*

AND TADASU OYAMA

*Hokkaido University*

Quite a few important works in experimental psychology, especially in the field of visual perception, have been done in Japan, and yet only a few of them are known to psychologists in the United States and Europe. Concerning figural aftereffects, Japanese psychologists have conducted a great many experiments since Gibson's (6), Köhler's (25), and Köhler and Wallach's (26) original studies were reported, though some related studies had been done before (33, 44). In the present paper, the authors intend to review some of these Japanese investigations, and hope to bring about some fruitful comparisons of American and European studies with them.

It will be worth while to outline briefly the Köhler-Wallach theory of figural aftereffects before reviewing these individual studies. If a part of the visual field has been occupied for some time by a figure, another figure which is afterwards shown in about the same place will generally be changed in its apparent location, size, shape, clearness, or depth. This phenomenon was named the "figural aftereffect." According to Köhler and Wallach, the most fundamental principle of the figural aftereffect is "displacement." The test-object or its parts recede from the region in which the inspection-object has been shown, and particularly from the place formerly occupied by the edge or contour of the object. If the T-object lies entirely within the area of the previously inspected figure, its

parts recede from the zone that has been occupied by the contour of the I-figure, and the T-object shrinks. Conversely, if the T-object surrounds the area of the I-object, the T-object is enlarged for the same reason. If parts of the T-object are displaced in varying degrees or in different directions, the shape of the T-object is distorted.

The second important principle of the Köhler-Wallach theory is "distance paradox." The amount of the displacement depends on the distance between the I- and T-object. For instance, if the I-object is a straight line, the T-line which coincides with the I-line will not be displaced. Neither will it be displaced if it is shown very far apart from the I-line. In a wide range of intermediate positions the T-line will recede from the I-line, and at a certain distance within this range its displacement will be maximal. Up to such an optimal distance, the farther the T-line is from the I-line, the larger is the displacement; this principle is called "distance paradox."

### GIBSON'S "CURVED LINE" EFFECT

More than twenty years ago, Gibson discovered a new phenomenon. If a person observes a slightly curved line continuously, it gradually comes to appear less curved. When a straight line is shown in the same place immediately afterwards, it appears curved in the opposite direction. This phenomenon was called "curved line" effect. More recently,

Köhler and Wallach proposed that this effect could be explained by the displacement of the test-line from the satiation area caused by the I-curve, and by the "distance paradox" principle of displacement. Since then, this "curved line" effect has been treated as a part of figural aftereffects.

In Japan, Nozawa (41) has performed the most systematic analysis

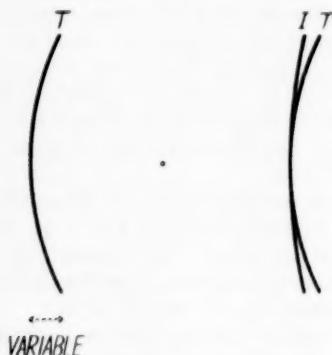


FIG. 1. INSPECTION- AND TEST-FIGURE USED BY NOZAWA (41)

of this effect. Gibson used a flexible rod, whose curvature could be varied, as a T-line to measure the amount of the effect. The observer was asked to adjust this flexible T-rod to appear straight, and the difference between this apparent straight line and the physically straight line was regarded as the amount of aftereffect. Nozawa, on the other hand, presented to the subject two T-lines on either side of a fixation mark; one line was straight or had a fixed curvature and was presented where the I-line had been, and the other was a flexible rod and was shown in the neutral area (see Fig. 1). His subject was asked to adjust the flexible rod to appear to have the same curvature as the fixed one. This new method made it possible for him to measure the aftereffect of the I-curve on T-curves with varied curva-

tures, as well as on the *straight T-line*. The results of this study of the extended "curved line" effect revealed the fact that the T-curve always decreased its apparent curvature, even when it was more curved than the I-curve. According to the Köhler-Wallach theory, the T-curve would have been expected to increase its apparent curvature. Nozawa criticized the theory on the basis of these and some other results. He also conducted experiments on other aspects of this effect. Some of them will be referred to later.

Yoshida (63) and Kogiso (23) measured the displacement of T-dots caused by the preceding inspection of curved lines. These are also analyses of the "curved line" effect in the broad sense, and will be mentioned later.

Oguro (49) studied Gibson's "tilted line" effect and obtained results which fitted the Köhler-Wallach theory well. Ishigooka (20) analyzed Gibson's "bent line" effect and found the maximal effect after the inspection of a right-angled figure.

#### KÖHLER AND WALLACH'S "SIZE" EFFECT

If we locate an outline circle as a T-object within another outline circle adopted as an I-object, the T-circle will shrink, and when it surrounds the I-circle, it will grow, according to the Köhler-Wallach "displacement" theory. Such shrinkage or growth will become maximal at a certain spatial separation between the outlines of I- and T-circles, according to the Köhler-Wallach "distance paradox" principle.

To examine these expectations, Oyama (52), Ikeda (13), Ikeda and Obonai (17), and Kogiso (24) measured the amount of growth and shrinkage of the T-circle by varying the size of the I-circle from smaller

to larger than the former. Some of these results are illustrated in Fig. 2, in which the abscissa represents the relative size of the I-circle to the T-circle, while the ordinate represents the relative amount of aftereffect. These curves, as well as those not quoted here, have essentially the same shape in spite of the variety of experimental conditions, such as the absolute size of the T-circle, the observation distance, the method of measurement, the inspection time, etc.

Oyama (55) discussed these results with the following conclusions:

a. The T-circle grows when it is larger than the I-circle and shrinks when it is smaller. This fact agrees exactly with the "displacement" principle.

b. If the size of the T-circle is equal to that of the I-circle and coincides with it, according to the "displacement" principle neither growth nor shrinkage should occur. On the contrary, Fig. 2 indicates that it

shrinks under such conditions. Köhler and Wallach previously recognized this fact and presented an additional hypothesis to meet it. However, Hebb (10) and Smith (58) objected that the additional hypothesis was an *ad hoc* one.

c. The amount of shrinkage, in general, is greater than the amount of growth. This is also underivable from the simple "displacement" principle.

d. There are, as the "distance paradox" principle predicts, optimal points of growth and of shrinkage, where the amount of growth or shrinkage becomes maximal. In almost all curves in Fig. 2, the maximal growth occurs when the I-circle is one-half of the T-circle in diameter, and the maximal shrinkage occurs when the I-circle is twice as large as the T-circle. This rule holds regardless of the size of the T-circle. It means that the optimal condition for displacement, or the limit of "distance paradox," is not determined by the absolute distance between the

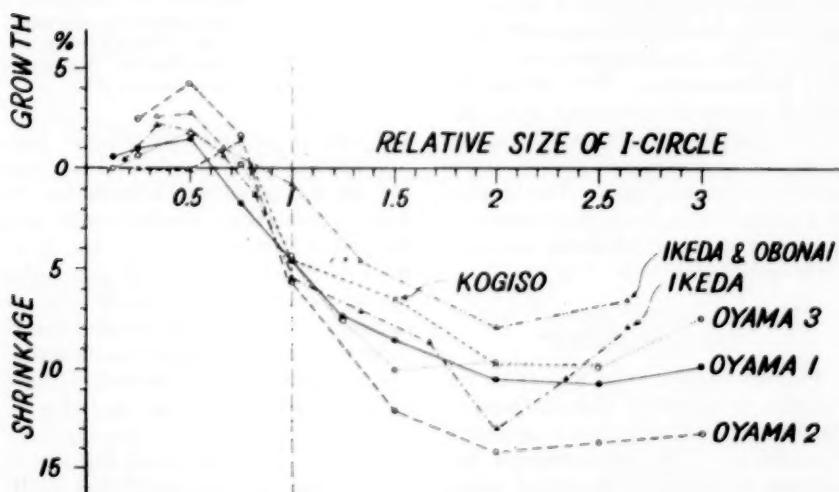


FIG. 2. AFTEREFFECT AS A FUNCTION OF THE SIZE RATIO OF THE INSPECTION-CIRCLE TO THE TEST-CIRCLE (SUMMARIZED FROM IKEDA [13], IKEDA AND OBONAI [17], KOGISO [24], AND OYAMA [52])

outlines of I- and T-circles, but by the relative size of the I-circle to the T-circle. Similar facts were discovered by Morinaga (33) and Ogasawara (47) with simultaneous assimilation-contrast illusion of concentric circles in which the maximal attraction effect occurred when the size ratio of the two circles was 2:3 or 3:2. In this illusion, the direction of "displacement" and the numerical values of the ratio at which the maximal effects are obtained are different from those of the figural aftereffects. In spite of these differences, the fact that the principle of ratio-relation applies adequately to both phenomena strongly suggests that there is a fundamental similarity between the two phenomena.

The close relationship between the figural aftereffects and the assimilation-contrast illusion was experimentally ascertained by Ikeda and Obonai (17). They discovered the continuous transition of results from one to the other of these two phenomena as shown in Fig. 3, as they varied the temporal condition of presentation of the two circles gradually from simultaneity to succession by means of a tachistoscope. The curves obtained under simultaneous presentation are similar to Ogasawara's, and those under successive presentations are like those in Fig. 2. The process of gradual shift from the former to the latter is observed under intermediate conditions.

#### "DISPLACEMENT" EFFECT AND "FIELD STRENGTH"

The displacement of a part of the T-object is the most essential part in the Köhler-Wallach theory of figural aftereffects. The fundamental hypothesis in Köhler's theory of visual perception is that a percept has a field of influence surrounding it, just as an electric charge has. Many ex-

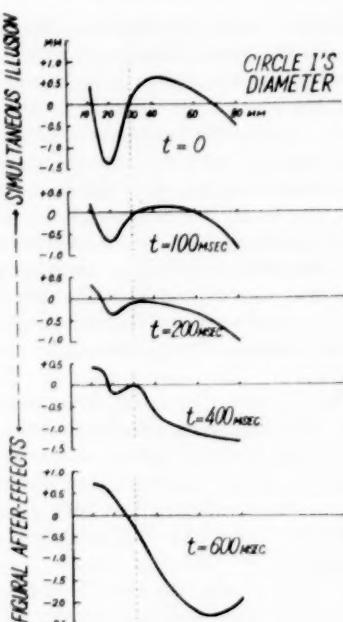


FIG. 3. AFTEREFFECT AS A FUNCTION OF THE SIZE OF INSPECTION-CIRCLE AT EACH STAGE OF ASYNCHRONISM OF PRESENTATION OF INSPECTION- AND TEST-FIGURE. EXPOSURE TIME OF BOTH INSPECTION- AND TEST-FIGURE IS 500 MSEC., AND THE START OF EXPOSURE OF T-FIGURE IS DELAYED ZERO TO 1,000 MSEC. FROM THAT OF I-FIGURE. THIS DELAY IS INDICATED BY  $t$  IN THIS FIGURE (REDRAWN FROM IKEDA AND OBONAI [17])

perimental analyses of visual phenomena have been conducted in Japan (36, 48, 62) on similar hypotheses. It was natural that Yoshida (63) and Kogiso (23) minimized the size of T-objects to small dots and measured the displacement of these T-dots located at various points around the I-object. Fig. 4 indicates the direction and amount of displacement of these dots. The results show that the T-dots not only recede from the area which has been occupied by the I-object, but also are attracted to it, and that the theoretical expectations from the ordinary figural aftereffects do not necessarily agree with the ex-

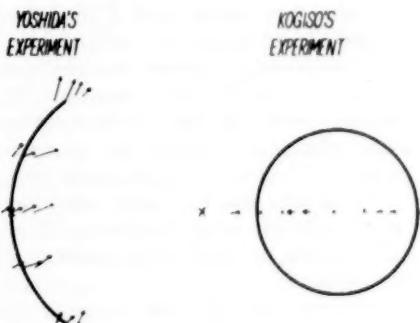


FIG. 4. YOSHIDA'S (63) AND KOGISO'S (23) EXPERIMENTS ON THE "DISPLACEMENT" EFFECT. ARROWS INDICATE THE DIRECTION AND AMOUNT OF DISPLACEMENT. THE LENGTHS OF ARROWS ARE DOUBLED RELATIVE TO INSPECTION-FIGURE

perimental facts in these situations.

Morinaga (34), Nakai (40), Oyama (54), and Ikuta (19) also measured the amount of displacement of the T-object. The results of these experiments as well as those of Fox (3) in the United States are shown in Table 1. Experimental conditions, methods of measurement, optimal distance between I- and T-objects, and the amount of maximal displacement in these experiments are compared.

Numbers in the "Method" column in this table indicate the kinds of measurement situations.

Oyama used dots both as I-objects and T-objects and measured the apparent growth and shrinkage of the distance between a pair of T-dots caused by a pair of I-dots. His results tell us that the amount of displacement does not depend upon the absolute distance between the I-dot and the T-dot, but rather upon the distance ratio of I-dots to T-dots, in the same manner as the "size" effect mentioned above.

For determination of the "field" strength at various points around the I-object, Nozawa (43) adopted as a measure the change of the stimulus threshold of a light spot. His I-objects were a curved line, a circle, and a tilted line. In general, these figures produced a sensitizing effect on one side of their lines and a desensitizing effect on the other side. He discovered that, in an ordinary experimental situation, displacement always occurred from the sensitized area to the desensitized area.

Motokawa, Nakagawa, and Kohata (38) studied figural aftereffects from

TABLE 1  
OPTIMAL DISTANCES BETWEEN INSPECTION- AND TEST-OBJECTS, AND AMOUNTS OF MAXIMAL DISPLACEMENT, DISCOVERED BY VARIOUS INVESTIGATORS

Investigator	Method	Observation Distance	Optimal Distance between I and T	Maximal Displacement
Fox (3), Exp. 1	1	203 cm.	1.3 cm. (22°)	1.3 mm. (2.2°)
Morinaga (34)	1	100 cm.	1.5 cm. (52°)	2.5 mm. (8.6°)
Oyama (54), Exp. 1, 2	1	300 cm.	1~4.5 cm. (10°~50°)	3.0 mm. (3.5°)
Oyama (54), Exp. 3, 4	1	61 cm.	0.5~3 cm. (0.5°~3°)	1.3 mm. (7.3°)
Oyama (54), Exp. 5	1	79 cm.	0.75~3 cm. (0.5°~2°)	1.3 mm. (5.7°)
Ikuta (19)	1	177 cm.	1.5 cm. (29°)	1.4 mm. (2.7°)
Köhler & Wallach (26)	2	274 cm.	0.6 cm. (8°)	2.2 mm. (2.8°)
Fox (3), Exp. 2	2	203 cm.	0.6~1.3 cm. (10°~22°)	0.8 mm. (1.3°)
Oyama (54), Exp. 6	2	61 cm.	0.5~1 cm. (0.5°~1°)	1.0 mm. (5.6°)
Oyama (54), Exp. 7	2	79 cm.	?	0.7 mm. (3.0°)
Yoshida (63)	3	70 cm.	?	2.3 mm. (11°)
Kogiso (23)	3	115 cm.	1~4 cm. (0.5°~2°)	9.3 mm. (25°) ?
Nakai (40)	4	130 cm.	1 cm. (26°)	1.1 mm. (2.9°)

the point of view of Motokawa's theory of retinal induction (35, 36, 37). Their method of measuring induction was essentially the same as those in many other psychophysiological studies of visual processes conducted by Motokawa and his collaborators which were reviewed by Gebhard (5). The electrical sensitivity of the dark-adapted eye after exposure to the light is compared with that of the eye at rest, and the change is regarded as an index of retinal induction. In their experiments with figural aftereffects, in addition to presenting a light-pattern which corresponded to the T-object, another light-pattern which corresponded to the I-object was presented, preceding the former. They proposed from their results that "displacement" and "distance paradox" could be explained by retinal induction.

#### TEMPORAL FACTORS

Concerning the temporal aspects of figural aftereffects, Gibson and Radner (7) reported with respect to the "tilted line" effect that the aftereffect increased with the increase in the duration of the inspection period, with the maximum effect at about 45

seconds; and Bales and Follansbee (2) found regarding the "curved line" effect that the aftereffect was greatest immediately after the inspection period and decreased within 60 seconds. More recently, Hammer (9), on "displacement" effect, and Nozawa (41), on "curved line" effect, obtained practically the same results on these two aspects in their more systematic experiments.

In these studies, the amount of aftereffect was measured by the method of adjustment, which required several seconds for one setting, and consequently the aftereffect several seconds after the inspection period was recorded. Oyama (51) pointed this out and tried to measure the aftereffect immediately after the inspection period by the method of constant stimuli. According to his results, one second of inspection is long enough to produce considerable amount of aftereffect, and a longer inspection period could hardly bring about any increase in the effect. He explained that the curves of disappearance of aftereffects started from almost the same level regardless of the inspection time, but the longer the inspection period was, the slower

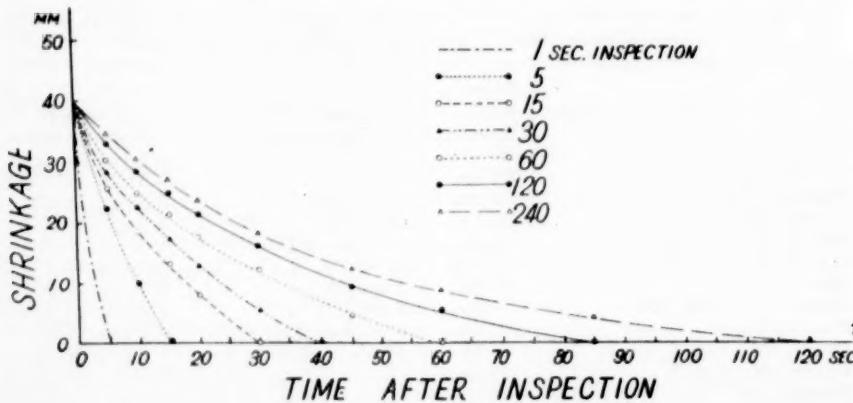


FIG. 5. COURSES OF DISAPPEARANCE OF AFTEREFFECTS AFTER INSPECTION-PERIODS OF VARIOUS DURATION (REDRAWN FROM IKEDA AND OBONAI (15))

was the rate of decrease; and that the four studies mentioned above were concerned with the lower points of the curves, while his own compared the curves at their starting points. He obtained such curves for various inspection periods. Ikeda and Obonai (15) performed essentially the same experiment more thoroughly and obtained similar results; Fig. 5 shows their results. The curves represent the course of disappearance of aftereffect for inspection periods of from 1 to 240 seconds respectively. The curves in Fig. 5 start from almost identical levels, and the longer the inspection period is, the slower is the rate of decrease. Oyama, as well as Ikeda and Obonai, proposed a mathematical formula for the disappearance of aftereffect on the basis of experimental data,

$$A = A_0 e^{-\lambda t}$$

in which  $A$  indicates the amount of aftereffect,  $A_0$  is a constant which represents the common starting point,  $\lambda$  is a parameter which represents the rate of decrease, and  $t$  is the time elapsed after the inspection period. This formula is essentially the same as that of Mueller (39), who used Hammer's data, except that  $\lambda$  is a parameter which varies as a function of inspection time, and not a constant as in Mueller's formula.

In addition to the previously mentioned two experiments, measurements of aftereffect immediately after inspection, as a function of inspection time, were conducted by Obonai and Suto (45), Fujiwara and Obonai (4), Suto and Ikeda (60), and Oyama (55). In three out of these six experiments, no significant difference was found between the amount of aftereffect of one-second inspection and that of 15 or 60 seconds' inspection; but, in the other experiments, the aftereffect somewhat increased as the

inspection time was lengthened. Why does such an inconsistency arise among the results of these experiments? The question cannot be answered yet.

As a general conclusion, we are able to say that even one-second inspection produces considerable amount of aftereffect, and longer inspection brings no or slight increase of aftereffect, but it slows down the rate of decrease markedly. In other words, the duration of inspection period affects the rate of decrease more strongly than it does the amount of aftereffect immediately after inspection.

Pertaining to the temporal factors, Nozawa (42) also conducted an interesting experiment in which he examined the effect of intermittent presentation of the I-figure. He adopted the same straight line of light both as an I-object and a T-object, and discovered that its apparent length decreased after the prolonged inspection of it. This aftereffect could also be produced by the intermittent presentation of the I-line, in greater amount than by the continuous presentation. Varying the cycle of on-and-off of I-line, the aftereffect became maximal when the light and the dark periods were one-half second each.

When, in the figural aftereffects experiment, the I-object has the same shape and size as the T-object, and both the inspection and test periods are shortened to a fraction of a second, the experimental situation becomes just the same as that of tau effect, which was studied by Scholz (57), Helson and King (11), and some other investigators. Tau effect was described as the overestimation and underestimation of the distance between the first stimulus and the second stimulus in a situation like that of the apparent-movement experi-

ment. Obonai and Suzumura (46) analyzed the relationship between ordinary figural aftereffect and tau effect, and concluded that these two kinds of phenomena were different phases of the same process. They observed a gradual transition from the figural aftereffect to the tau effect. When the exposure time of the first stimulus was short, both stimuli were displaced, and when it was rather long, only the second stimulus was displaced. Usually, the former is called tau effect and the latter, figural aftereffect. It was also discovered that in the figural-aftereffect situation, not only the repulsion of the second stimulus by the first stimulus but also the attraction of the second by the first occurred, as it did in the tau-effect situation. This fact contradicts the Köhler-Wallach displacement theory.

#### SOME OTHER SPATIAL FACTORS

According to the "localization" of aftereffect discussed by Gibson (6) and Köhler (25), a T-figure presented at a different place from the I-figure should not be affected, or should have a weak aftereffect, if any. Nozawa (41) measured the amount of "curved line" effect, varying the spatial relationship between the I-curve and the T-line from overlap to separation. The results were not simple. Maximal amount of aftereffect occurred when the T-line was presented a little inside or outside of the I-curve, and not when they were overlapping. Oyama (52, 55) reported, concerning the "size" effect, that a considerable amount of aftereffect was found when the I-circle was presented eccentrically to the T-circle, and that the aftereffect slightly decreased as the distance between the centers of I- and T-circles increased, but it occurred even when there was no overlapping of the circles.

When the I-figure and the T-figure are presented in different distances from the subject, the following question arises. Which is the determining factor of aftereffect, the retinal size or the apparent size of the I-circle? To answer this question, Oyama performed some experiments, using various sizes of I-circles which were at different observation distances from the T-circle. His results were not conclusive, but it was certain that the retinal size of the I-circle was, at least, a determining variable for aftereffect, although it was possible that the apparent size was also a simultaneous determining factor, as Sutherland (59) discussed.

Concerning the relation between figural aftereffects and the laws of organization in the Gestalt school, Mori and Nagashima (32) performed some analyses. They investigated the effect of the tridimensional organization of T-figure on the aftereffect. Like Luchins and Luchins (29), they used as T-figures a complete and several incomplete Necker cubes, and measured the change in size of their square parts. His results showed that a tridimensional organization in a T-figure depressed the "size" effect on the T-figure in the cases of a Necker cube and other cubic figures, but such depressive effect was not found in noncubic figures in his control experiments. Mori (31) analyzed the effect of closure, using triangles and circles with various degrees of closure as I-figures. The results were not so clear-cut, but it could be argued that the discontinuous decrease of aftereffect corresponded to the abrupt change in the characteristics of the seen figure.

When two I-objects are presented simultaneously, how do their after-effects interact? To answer this question, Ikeda (14) and Oyama (53) performed some experimental analyses.

In Ikeda's experiment, aftereffects of two I-circles upon one T-circle were examined; and in Oyama's experiment, aftereffects of two I-circles upon two T-circles were analyzed. Their results suggested algebraic summation of aftereffects.

On the effect of the width of the outline of the I-circle, Oyama (52, 55) conducted some experiments and found that "size" effect slightly increased as the width increased, while Graham (8) did not find any influence of the width of the I-line in her experiments on displacement.

#### LUMINANCE, CONTRAST, ILLUMINANCE, AND COLOR

As to the influence of the luminance of the I-figure upon its aftereffect, Fujiwara and Obonai (4) found that the amount of "size" effect increased with the increase in the luminance of the I-figure when a luminous stimulus in a dark room was used as the I-figure. Yoshida (64) also obtained the same result in his experiment in which a gray figure with various reflectance on a black background was adopted as the I-figure.

As to the contrast of a gray I-figure to a white background (the luminance difference between the I-figure and the background), Fujiwara and Obonai could not find its influence, but Yoshida reported that the aftereffect increased as the contrast increased and became maximum at the highest degree of contrast. The latter agrees with Graham's results on the displacement effect. Nozawa, in his experiment on "curved line" effect, varied not only the contrast of the I-figure but also that of the T-figure, and discovered that the aftereffect increased as the contrast of the I-figure increased, but decreased as the contrast of the T-figure increased. It was also found by Nozawa that the aftereffect grew more rapidly during the

inspection period when the contrast of the I-figure was greater.

When the intensity of illumination was changed and the illuminance of stimulus field was varied, keeping the relative contrast of I- and T-figures to their background constant, no change in aftereffect could be found in Fujiwara and Obonai's experiment. This result agrees with Graham's results.

Takagi and Ishikawa (61) used chromatic stimuli as I- and T-objects and analyzed the effect of color upon figural aftereffects. In their results, when the colors of I- and T-objects were the same (for instance, they were both red or green), considerable amount of aftereffect was observed; but when their colors were different (for instance, the I-object was red and T-object was green), the aftereffect was not so conspicuous.

#### SOME RELATED EXPERIMENTS

Gibson's "adaptation" and Köhler and Wallach's "self-satiation," which occur during the prolonged inspection of a figure, may be regarded as the aftereffect of the I-figure on the T-figure which is identical with the I-figure. Nozawa (41) and Ikeda and Obonai (16) performed experimental studies of these phenomena from such a point of view. Nozawa found that the adaptation to the curved line became more prominent as its curvature increased. Ikeda and Obonai discovered that, by prolonged inspection, a circle came to appear smaller, a curve less curved, and the distance between two parallel lines became narrower and their length shorter. These effects increased rapidly at the beginning and then increased rather gradually.

During the prolonged inspection of a figure, some other phenomena are also observed, as Marks (30) has pointed out. Sakurabayashi (56) in-

dependently discovered this fact. His subjects observed, during the prolonged inspection of some rather complex figures, that the typical sight, which appeared at the first stage of inspection and obeyed Gestalt laws of visual organization, was lost, and nontypical and irregular sights appeared one after another as the inspection period was prolonged. Oyama's and Ikeda and Obonai's subjects, like Marks' subjects, also reported that the I-circle came to appear like a polygon rather than a circle, to be distorted, to make an auto-kinetic movement, or to lose a part of its outline.

The fact that the repetition of adjustment of a Müller-Lyer figure decreases the amount of illusion has been known as the "practice effect" since Judd found it at the beginning of this century. Recently, Köhler and Fishback (27, 28) asserted that this fact could be explained by "satiation" just as figural aftereffects could be. Azuma (1) examined this problem and concluded from his experimental results that the careful observation of every part of the illusion figure was as effective in decreasing the amount of illusion as the repeated adjustment, but that the presence of "the pattern of satiation," mentioned in the Köhler-Fishback theory, might not always be the necessary condition nor the satisfactory condition for bringing the effect.

If, prior to the observation of the figure-ground reversible patterns or the reversible perspective patterns, the figure which corresponds to one

of the two shapes involved in these patterns has been inspected for a long time, the other shape becomes dominant for a while. This kind of aftereffect was discovered by Hochberg (12) and Oyama (50) independently. Kakizaki (21, 22) found a similar effect in the binocular rivalry, and suggested the importance of figure-relationship besides the eye-relationship between the preceding situation and the test situation.

### SUMMARY

Many experimental studies on figural aftereffects in Japan were reviewed under several topics: Gibson's "curved line" effect; Köhler and Wallach's "size" effect; "displacement" effect and "field strength"; temporal factors; some other spatial factors; luminance, contrast, illuminance, and color; and some related experiments. Concerning some typical experimental situations, the effects of spatial and temporal variables were analyzed quantitatively, and some mathematical functions to relate these variables to the amount of aftereffects were examined. Size ratio of the I-object to the T-object, the duration of inspection period, and the time interval between inspection and test period were found to be important parameters in figural aftereffects. It was indicated that the assimilation-contrast illusion and the tau effect had close kinship to figural aftereffects. The influence of some other stimulus factors and experiments on some related phenomena were also reviewed.

### REFERENCES

1. AZUMA, H. The effect of experience on the amount of the Müller-Lyer illusion. *Jap. J. Psychol.*, 1952, **22**, 111-123.\*
2. BALES, J. F., & FOLLANSBEE, G. L. The after-effect of the perception of curved lines. *J. exp. Psychol.*, 1935, **18**, 499-503.
3. FOX, H. B. Figural after-effects: "satiation" and adaptation. *J. exp. Psychol.*, 1951, **42**, 317-326.
4. FUJIWARA, K., & OBONAI, T. The quantitative analysis of figural after-effects: II. Effects of inspection time and the intensity of light stimulus upon the

amount of figural after-effects. *Jap. J. Psychol.*, 1953, **24**, 114-120.\*

5. GEBHARD, J. W. Motokawa's studies on electric excitation of the human eye. *Psychol. Bull.*, 1953, **50**, 73-111.
6. GIBSON, J. J. Adaptation, after-effect and contrast in the perception of curved lines. *J. exp. Psychol.*, 1933, **16**, 1-31.
7. GIBSON, J. J., & RADNER, M. Adaptation, and after-effect in the perception of tilted lines: I. Quantitative studies. *J. exp. Psychol.*, 1937, **20**, 453-467.
8. GRAHAM, ELAINE H. Figural after-effects as a function of contrast, area and luminance. (Unpublished. Quoted in Prof. C. H. Graham's seminar in Kyoto, 1952.)
9. HAMMER, E. R. Temporal factors in figural after-effects. *Amer. J. Psychol.*, 1949, **62**, 337-354.
10. HEBB, D. O. *The organization of behavior*. New York: Wiley, 1949.
11. HELSON, H., & KING, S. M. The tau effect: An example of psychological relativity. *J. exp. Psychol.*, 1931, **14**, 202-217.
12. HOCHBERG, J. E. Figure-ground reversal as a function of visual satiation. *J. exp. Psychol.*, 1950, **40**, 682-686.
13. IKEDA, H. Studies in figural after-effects. Tokyo Bunrika University Dissertation, 1951, (unpublished).\*\*
14. IKEDA, H. On the cancellation of two opposing figural after-effects. *Jap. J. Psychol.*, 1956, **26**, 407-410.\*\*
15. IKEDA, H., & OBONAI, T. The quantitative analysis of figural after-effects: I. The process of growth and decay of figural after-effects. *Jap. J. Psychol.*, 1953, **23**, 246-260; **24**, 59-66.\*
16. IKEDA, H., & OBONAI, T. The quantitative analysis of figural after-effects: II. On "self-satiation." *Jap. J. Psychol.*, 1953, **24**, 179-192.\*
17. IKEDA, H., & OBONAI, T. The studies in figural after-effects: IV. The contrast-confluence illusion of concentric circles and the figural after-effect. *Jap. psychol. Res.*, 1955, **2**, 17-23.
18. IKEDA, H., & OBONAI, T. Figural after-effect, retroactive effect and simultaneous illusion. *Jap. J. Psychol.*, 1955, **26**, 235-246.\*
19. IKUTA, H. Displacement in figural after-effects and simultaneous-illusions. *Jap. J. Psychol.*, 1956, **27**, 218-226.\*
20. ISHIGOOKA, Y. Studies of figural after-effects (the second report). Reported at the 20th Annual Meeting of J. P. A., 1956.
21. KAKIZAKI, S. The effects of preceding conditions upon binocular rivalry: (I). *Jap. J. Psychol.*, 1950, **20**, No. 2, 24-32.\*
22. KAKIZAKI, S. The effects of preceding conditions upon binocular rivalry: (II). *Jap. J. Psychol.*, 1950, **20**, No. 4, 11-17.\*
23. KOGISO, I. An experiment on the displacement in figural after-effects. *Jap. J. Psychol.*, 1956, **26**, 405-407.\*\*
24. KOGISO, I. (unpublished data.)
25. KÖHLER, W. *Dynamics in Psychology*. New York: Liveright, 1940.
26. KÖHLER, W., & WALLACH, H. Figural after-effects. An investigation of visual process. *Proc. Amer. phil. Soc.*, 1944, **88**, 269-357.
27. KÖHLER, W., & FISHBACK, J. The destruction of the Müller-Lyer illusions in repeated trials: I. An examination of two theories. *J. exp. Psychol.*, 1950, **40**, 267-281.
28. KÖHLER, W., & FISHBACK, J. The destruction of the Müller-Lyer illusions in repeated trials: II. Satiation patterns and memory traces. *J. exp. Psychol.*, 1950, **40**, 398-410.
29. LUCHINS, S., & LUCHINS, H. On the relationship between figural after-effects and the principles of Prägnanz. *Amer. J. Psychol.*, 1952, **65**, 16-26.
30. MARKS, M. R. Some phenomena attendant on long fixation. *Amer. J. Psychol.*, 1949, **62**, 392-398.
31. MORI, T. Figural after-effects and stimulus-organization. *Jap. psychol. Res.*, 1956, **3**, 8-14.
32. MORI, T., & NAGASHIMA, K. The effects of the organization of the total patterns upon figural after-effects. *Essays and Studies by Members of Tokyo Woman's Christian College*, 1953, 87-107.\*
33. MORINAGA, S. Conditions of size-assimilation and size-contrast. *Masuda Hakushū Shaon Saikin Shinrigaku Ronbunshu*, 1935, 28-48.\*\*
34. MORINAGA, S. Optical illusions and figural after-effects. Reported at the 19th Annual Meeting of J. P. A., 1955.
35. MOTOKAWA, K. Physiological induction in human retina as basis of color and brightness contrast. *J. Neurophysiol.*, 1949, **12**, 475-488.
36. MOTOKAWA, K. Field of retinal induction and optical illusion. *J. Neurophysiol.*, 1950, **13**, 413-426.
37. MOTOKAWA, K. Retinal traces and visual perception of movement. *J. exp. Psychol.*, 1953, **45**, 369-377.

\* Japanese text with English summary.

\*\* Japanese text without English summary.

38. MOTOKAWA, K., NAKAGAWA, D., & KOHATA, T. Figural after-effects and retinal induction. *J. gen. Psychol.*, (in press).
39. MUELLER, C. G. Numerical transformations in the analysis of experimental data. *Psychol. Bull.*, 1949, **46**, 198-223.
40. NAKAI, K. (unpublished data.)
41. NOZAWA, S. Prolonged inspection of a figure and the after-effect there-of. *Jap. J. Psychol.*, 1953, **23**, 217-234; **24**, 47-58.\*
42. NOZAWA, S. On the after-effect by intermittent presentation of inspection figure. *Jap. psychol. Res.*, 1955, **2**, 9-16.
43. NOZAWA, S. An experimental study on figural after-effect by the measurement of field strength. *Jap. psychol. Res.*, 1956, **3**, 15-24.
44. OBONAI, T., & SATO, Y. Contributions to the study of psycho-physiological induction: (11) The study of successive contrast-assimilation. *Jap. J. Psychol.*, 1937, **12**, 451-464.\*
45. OBONAI, T., & SATO, Y. Studies of figural after-effects by the inspection of short period. *Jap. J. Psychol.*, 1952, **22**, 248. (Abstract)\*\*
46. OBONAI, T., & SUZUMURA, K. Contributions to the study of psycho-physiological induction: (43) The characteristics of successive induction during periods immediately following retinal stimulation. *Jap. psychol. Res.*, 1954, **1**, 45-54.
47. OGASAWARA, J. Displacement-effect of concentric circles. *Jap. J. Psychol.*, 1952, **22**, 224-234.\*
48. OGASAWARA, J. "Field" structures in visual process. *Chiwa Sensei Kanreki Kinen Ronbunshu*, 1952, 3-8.\*
49. OGUCHI, S. After-effects of vertical and tilted lines. *Jap. J. Psychol.*, 1957, **27**, 303-305.\*\*
50. OYAMA, T. Figural after-effects and reversible figures. *Chiwa Sensei Kanreki Kinen Ronbunshu*, 1952, 47-55.\*\*
51. OYAMA, T. Experimental studies of figural after-effects: I. Temporal factors. *Jap. J. Psychol.*, 1953, **23**, 239-245.\*
52. OYAMA, T. Experimental studies of figural after-effects: II. Spatial factors. *Jap. J. Psychol.*, 1954, **25**, 195-206.\*
53. OYAMA, T. After-effects of two inspection figures. *Jap. J. Psychol.*, 1955, **26**, 202-203.\*\*
54. OYAMA, T. Experimental studies of figural after-effects: III. Displacement effect. *Jap. J. Psychol.*, 1956, **26**, 365-375.\*
55. OYAMA, T. Temporal and spatial factors in figural after-effects. *Jap. psychol. Res.*, 1956, **3**, 25-36.
56. SAKURABAYASHI, H. Studies in creation: IV. The meaning of prolonged inspection from the standpoint of creation. *Jap. J. Psychol.*, 1953, **23**, 207-216.\*
57. SCHOLZ, W. Experimentelle Untersuchungen über die phänomenale Grösse von Raumstrecken, die durch Sukzessiv-Darbietung zweier Reize begrenzt werden. *Psychol. Forsch.*, 1924, **5**, 219-272.
58. SMITH, K. R. The statistical theory of figural after-effect. *Psychol. Rev.*, 1952, **59**, 401-402.
59. SUTHERLAND, N. S. Figural after-effects, retinal size and apparent size. *Quart. J. exp. Psychol.*, 1954, **6**, 35-44.
60. SUTO, Y., & IKEDA, H. An examination of the relationships between the inspection time and the figural after-effects. *Jap. J. Psychol.*, 1957, **27**, 377-380.\*\*
61. TAKAGI, K., & ISHIKAWA, M. Experimental studies on after-effect: (1). Reported at the 20th Annual Meeting of J. P. A., 1956.
62. YOKOSE, Z. The law of the "field" in visual form perception: (1) A theoretical formula to seek the field strength of the form and its experimental proof. *Jap. psychol. Res.*, 1954, **1**, 55-64.
63. YOSHIDA, T. An experimental study of figural after-effect. *Jap. J. Psychol.*, 1953, **23**, 235-238.\*
64. YOSHIDA, T. Figural after-effects as a function of the brightness of figure and ground. Reported at the 19th Annual Meeting of J. P. A., 1955.

Received March 11, 1957.

## STIMULUS PREDIFFERENTIATION: SOME GENERALIZATIONS AND HYPOTHESES<sup>1,2</sup>

MALCOLM D. ARNOULT

*Air Force Personnel and Training Research Center*

At one time it was an accepted dictum in the field of verbal learning that attaching a new response to an old stimulus, according to the A-B . . . A-K paradigm, would lead to negative transfer. About 15 years ago a number of experiments began to be reported in which the same paradigm was used to produce positive transfer. The main characteristic of these new studies was the fact that the two sets of responses were sufficiently different that there was essentially no generalization between them: neither incompatibility nor facilitation. It was hypothesized that the pretraining "predifferentiated" the stimuli so that they were more "distinctive," or less "confusing." In recent years a substantial number of experiments have been devoted to this problem of stimulus predifferentiation; many potentially relevant variables have been investigated, some methodological improvements have been suggested and incorporated into later studies, and a number of hypotheses have been offered to account for the positive transfer obtained under these conditions.

Recently the writer surveyed a number of articles and dissertations on the topic of stimulus prediffer-

entiation in an attempt to discover what generalizations could be made from frequently conflicting results, and what evidence could be found for or against the various explanatory hypotheses. The results of this effort are incorporated in the first two sections of this paper. In the third section some suggestions are made concerning additional variables, the consideration of which might contribute to clearing up some of the ambiguities which presently exist. It should be emphasized that an attempt was made to limit the survey to those studies which conformed strictly to the stimulus predifferentiation paradigm, and it is believed that the survey is fairly complete within that area. Likewise, in the interests of brevity, the writer has sternly resisted the temptation to discuss the various issues in the larger context of transfer of training, although in many cases such an extension would be relevant. For these reasons it is hoped that the reader will use the conclusions and recommendations of this paper as a guide to the literature rather than as a substitute for it.

### GENERALIZATIONS

The survey indicated that there is enough agreement in results among the various experiments to provide generalizable conclusions in two broad areas. One of these has to do with the kind of verbal pretraining given, and the other concerns the amount of such training.

The categories of pretraining and the results achieved with each kind are summarized below, with examples of each category given in Table 1.

<sup>1</sup> This report is based on work done under ARDC Project No. 7706, Task No. 27001, in support of the research and development program of the Air Force Personnel and Training Research Center, Lackland Air Force Base, Texas. Permission is granted for reproduction, translation, publication, use, and disposal in whole or in part by or for the United States Government.

<sup>2</sup> Thanks are due to Dr. J. M. Vanderplas and Dr. Harold W. Hake for reviewing the manuscript.

TABLE 1

EXAMPLES OF POSSIBLE S-R PAIRS USED DURING DIFFERENT KINDS OF PREDIFFERENTIATION TRAINING WHEN THE TRANSFER TASK INVOLVES MOVING A CONTROL UPWARD IN RESPONSE TO A RED LIGHT AND DOWNWARD IN RESPONSE TO A GREEN LIGHT

Kind of Pretraining	Pretraining		Transfer Task	
	Stimulus	Verbal Response	Stimulus	Motor Response
Relevant S-R	Red light	"Up"	Red light	Up
	Green light	"Down"	Green light	Down
Relevant S	Red light	"Cow"		Same as above
	Green light	"Horse"		
Irrelevant S	Bright light	"Cow"		Same as above
	Dim light	"Horse"		
Attention	Red light	None		Same as above
	Green light	None		
No pretraining	None		Same as above	

The terminology is, in part, that suggested by McAllister (28).

#### *Categories of Pretraining*

*Relevant S-R.* In this type of pretraining the stimuli used for the pretraining task are identical to those used in the transfer task, and the responses used in the pretraining task are somehow symbolic of, or bear a sign-significate relation to, the responses used in the transfer task. Strictly speaking, this type of pretraining is used in the classical "transfer of training" studies and should not be considered an example of stimulus predifferentiation. Stimulus predifferentiation is usually characterized by the fact that the responses used in pretraining and in the transfer task are completely independent of one another, whereas in Relevant S-R training the kind of transfer obtained depends greatly upon the re-

lationship which exists between the two sets of responses. This type of pretraining has been used, however, in studies which derived their rationale from the predifferentiation hypothesis (7, 10, 12, 28). In those cases in which it has been used, Relevant S-R training has proved to be equal or superior to any other kind of verbal pretraining.

*Relevant S.* In this type of pretraining the stimuli used are identical to the ones used in the transfer task, but the responses are completely different from those used in the transfer task (1, 3, 8, 10, 11, 12, 13, 14, 15, 16, 17, 21, 25, 28, 34, 36). This is the kind of pretraining specified, for example, in the predifferentiation hypotheses of Gibson (18) and of Miller and Dollard (29). In most of the studies in which Relevant-S pretraining was compared with some other type of pretraining, the experi-

mental group performed better on the transfer task than a group given no pretraining or a group given any other kind of pretraining, except directed attention (see below). Battig, however, has shown that the effectiveness of Relevant-S training decreases as task complexity increases (10).

*Irrelevant S.* This type of pretraining is most often given in order to obtain a control group having the same "performance set" as the experimental group. The stimuli used in the pretraining task are different from those used in the transfer task but are equated with them in difficulty. In none of the experiments surveyed was the performance of the Irrelevant-S group on the transfer task superior to the performances of groups given training in attention or given no pretraining at all (1, 8, 13, 14, 15, 17, 28).

*Attention.* In this type of training, *S* is not required to make any sort of overt differential responses to the stimuli during the pretraining period, in the sense of learning "labels" for them. He is required, however, by instructions or some other means to attend to the distinctive characteristics of the stimuli. This type of training was consistently superior to Irrelevant-S training. It was as effective as Relevant-S training in 50 per cent of cases in which they could be directly compared (3, 11, 12, 13, 21, 25, 34, 36).

*No pretraining.* This group starts on the transfer task without any previous experience in the experimental situation. In general, this type of control group is unsatisfactory in that there is no control for the factors of performance set or attention.

A summary of the results obtained in experiments in which various kinds of predifferentiation training were

given indicates that the following generalizations may be made: (a) *Relevant S-R* training (if it can be accepted as falling into the category of stimulus predifferentiation) is the most effective form of verbal pretraining; (b) *Relevant-S* pretraining is, in most cases, more effective than any pretraining method except *Relevant S-R* training; (c) *Irrelevant-S*, or performance set, pretraining is usually poorer than other pretraining methods; and (d) *Directed Attention* pretraining is often as effective as *Relevant-S* pretraining.

#### *Amount of Predifferentiation Training*

Of the studies included in the survey, the following were concerned with varying the amount of predifferentiation training: Arnoult (3), Baker and Wylie (7), Baldwin (8), J. H. Cantor (14), Gagné and Baker (16), Goss (21), and Rossman and Goss (35). The results of these experiments are summarized in Table 2. In order to facilitate comparisons among the experiments, the number of pretraining trials reported has been expressed as the number of experiences *S* had with each stimulus.

These results may be summarized, perhaps, by the following generalization: *Positive transfer from stimulus predifferentiation training may be expected after a minimum of 4 to 8 pre-training trials and reaches a maximum after 8 to 12 pre-training trials.* Support for this generalization may be found in the recent experiment by Arnoult (3). Using a transfer task (shape recognition) which was quite different from those used in the other experiments cited above, he measured two levels of achievement on the transfer task as a function of the number of pretraining trials. The curves obtained were monotonic, negatively accelerated functions which

TABLE 2  
SUMMARY OF EXPERIMENTS IN WHICH AMOUNT OF PRETRAINING WAS VARIED

Experiment	Comparison	Greater Number of Trials Produced More Positive Transfer
Baker and Wylie (7)	2 trials vs. 0 trials 8 trials vs. 0 trials	No Yes
Gagné and Baker (16)	2 trials vs. 0 trials 4 trials vs. 0 trials 8 trials vs. 0 trials	No Sometimes Yes
Rossman and Goss (35)	4 trials vs. 1 trial	No
Arnoult (3)	(5-15) trials vs. (1-4 trials)	Yes
Baldwin (8)	24 trials vs. 6 trials	No
Goss (21)	20 (avg.) trials vs. 12 (avg.) trials	No
Cantor (14)	24 trials vs. 12 trials 72 trials vs. 12 trials	No

rose rapidly and tended to level off in the vicinity of 8 to 10 pretraining trials.

#### HYPOTHESES

The next step in surveying the data on stimulus predifferentiation was to examine the implications of the results summarized above with respect to the various hypotheses which have been offered to account for the positive transfer obtained from predifferentiation training.

These hypotheses can be grouped more or less adequately into five categories which have been labeled as follows: (a) acquired distinctiveness of cues; (b) reduction in intralist generalization; (c) increased meaningfulness; (d) attention to cues; and (e) performance set. Each of these hypotheses was examined in turn in an attempt to weigh the evidence for and against it. It should be remembered that any experiment in which positive transfer was obtained can be considered as supporting whatever hypothesis the experimenter used as a starting point. Consequently, only those studies were considered in which two or more hypotheses were

directly compared, or in which a test was made of a deduction from one of the hypotheses.

*Acquired distinctiveness of cues.* This hypothesis was formulated by Miller and Dollard (29) and states that

... learning to respond with highly distinctive names to similar stimulus situations should tend to lessen the generalization of other responses from one of these situations to another since the stimuli produced by responding with the distinctive name will tend to increase the difference in the stimulus patterns of the two situations (30, p. 174).

The increase in differentiation which results from this process is called the acquired distinctiveness of cues. Miller and Dollard further hypothesize that removal of the verbal responses by repression "... will remove the basis for acquired distinctiveness and increase the amount of primary generalization" (30, p. 174).

The first part of the Miller-Dollard hypothesis implies that the amount of positive transfer resulting from verbal predifferentiation training will be a function of the degree to which the stimulus items are clearly differentiated during that training. One way in which the degree of differenti-

ation may be varied is by varying the specificity of the verbal labels attached to the stimuli. Hake and Eriksen (22) required their Ss to learn 2, 4, or 8 labels for 16 different stimuli. Following this training they were required to relearn to discriminate among the stimuli using 2, 4, or 8 new labels. All possible combinations of these conditions were investigated. The results showed that the specificity of the labels affected the speed of learning in both the pre-training and the transfer task, but that the speed of learning in the transfer task was independent of the number of labels which had been used during pretraining. Likewise, Robinson (34) found that the specificity of the labels used during paired-associate pretraining did not affect performance on a transfer task which required *S* to make same-different discriminations among the stimuli; and, in a second experiment, Hake and Eriksen found that label specificity did not affect subsequent recognition of forms (23). Thus it would appear that, while the Miller-Dollard hypothesis of acquired distinctiveness of cues seems eminently reasonable, the only available tests of a deduction from the hypothesis fail completely to support it.

Only one of the experiments covered by this survey included an attempt to test that part of the Miller-Dollard hypothesis dealing with the effects of repression. Rossman and Goss (35) had Ss learn to associate nonsense shapes with nonsense syllables to a criterion of mastery. One group was then given a single postcriterion trial on which electric shock was administered with every response. It was assumed that this noxious accompaniment to the response would lead to repression of the newly acquired associations. When these Ss were compared on the

transfer task with Ss who had received a normal postcriterion trial, no difference in performance was detected. The extent to which these results can be considered as harmful to the Miller-Dollard repression hypothesis depends, of course, on the extent to which one is willing to assume that the noxious stimulus was adequate for the purpose.

*Reduction in intralist generalization.* In 1940, E. J. Gibson formulated the following hypothesis (18, p. 222): "If differentiation has been set up within a list, less generalization will occur in learning a new list which includes the same stimulus items paired with different responses; and the trials required to learn the new list will tend to be reduced by a reduction of the internal generalization." The amount of the predicted transfer would be maximally positive in the case where there was a minimum of interlist response generalization. This hypothesis is very similar to the Miller-Dollard hypothesis and would lead to the same predictions with regard to the transfer of predifferentiation training. Gibson further hypothesized, however, that "Generalization will increase to a maximum or peak during the early stages of practice with a list, after which it will decrease as practice is continued" (18, p. 206). A subsequent experiment produced results confirming this prediction (19). It can be argued, then, that if paired-associate pretraining produces an initial increase in the tendency to confuse the stimulus items, followed by a decrease, then the transfer of such training should be negative after just a few trials, then positive as the number of pretraining trials increases. Gagné and Baker (16) found no difference in the amount of transfer after two or four pretraining trials, and Rossman and Goss (35) found one and four trials

to be equivalent with respect to the amount of transfer obtained (see Table 2). It should be pointed out that in the Gagné and Baker experiment neither of these experimental groups was consistently superior to a group receiving no pretraining at all, whereas in the Rossman and Goss experiment a zero-practice control group was not included. As was the case with the Miller-Dollard hypothesis discussed earlier, it would appear that, while the primary predifferentiation hypothesis has been supported by many experiments, attempts to test a specific deduction from the main hypothesis have yielded negative results.

*Increased meaningfulness.* Some writers have considered the possibility that the positive transfer from predifferentiation training is primarily due to an increase in meaningfulness of the stimuli as a result of having new responses associated with them. Arnoult has suggested, for example, the following:

If meaningfulness is measured in terms of the number of independent associations linked with a stimulus item, it would seem reasonable that adding a new association to a particular stimulus (by predifferentiation training) should make subsequent learning of that stimulus item easier (1, p. 402).

This idea is consistent with Noble's suggestion that:

... the procedure of endowing stimuli with the properties of meaningfulness (*m*) or familiarity (*f*) may constitute one unambiguous definition of Thorndikean "identifiability," which in turn may be related to such current notions as "predifferentiated structure," "distinctiveness," "cue-value," and "recognizability" (32, pp. 96-97).

Essentially the same idea has also been expressed by Dysinger (15). Some support for this hypothesis can be found in the experiments surveyed. Arnoult (3) found that in at least a limited way the amount of positive transfer to a recognition task was a

function of the meaningfulness of the response terms used during paired-associate pretraining. McAllister (28), likewise, found more positive transfer resulting from the use of some sorts of response terms than others, and the difference may have been due to differences in the meaningfulness of the response terms. On the other hand, Campbell and Freeman (12) found no relation between the meaningfulness of the responses used during pretraining and performance on a subsequent recognition test. A closer examination of the Arnoult and McAllister experiments suggests, furthermore, that the increase in meaningfulness occurring in both these experiments was not a function solely of the "number of associations" possessed by the response terms themselves, but rather was due to the introduction of a factor which might be called "belongingness" (37), i.e., the introduction of a pre-experimentally learned relationship between a pair of terms.<sup>3</sup> In the first case (Arnoult) this type of relationship existed between the stimuli and the responses used during pretraining, and in the second case (McAllister) it existed between the two sets of responses. While it is not unreasonable that increasing the meaningfulness of the response terms used during pretraining might enhance the effect of verbal pretraining, it appears doubtful that this factor alone will account for all of the positive transfer obtained from predifferentiation training.

*Attention to cues.* The results ob-

<sup>3</sup> For example, consider the four responses *kitchen*, *furry*, *Persian*, and *feline*. In terms of conventional measures of "meaningfulness" (31), these terms (considered by themselves) are probably arranged in descending order, i.e., they would elicit successively fewer associations. They are in ascending order, however, in terms of their "belongingness" to the stimulus word *cat*.

tained by Hake and Eriksen in the discrimination learning study in which response-specificity was varied led them to conclude:

The results appear to us to emphasize the importance of the general *labeling* process rather than factors related to the particular labels used. We may judge from our results and from others that the perceptual gain resulting from labeling practice appears to occur as long as *Ss* have a decision to make about the stimuli on each trial. The labeling task given to our *Ss* seems merely to have provided a context which defined an objective for them (22, pp. 166-167).

Similarly, Robinson suggested that the critical features of the pretraining were (a) attention to the stimuli and (b) active search for identifying features. He concluded that his results demonstrated that ". . . the learning of the arbitrary names for the . . . [stimuli] . . . did not produce any further change in stimulus discriminability" (34, p. 114). Those experiments which have included specific training methods based on directed attention to critical cues have, however, yielded ambiguous results. Campbell and Freeman (12), Robinson (34), and Smith and Goss (36) found this type of training to be as effective as standard verbal paired-associate training. G. N. Cantor (13) and Goss (21), on the other hand, found it to be less effective. Arnoult (3) and Kurtz and Hovland (25) obtained ambiguous results. In none of these experiments, however, was an attempt made to discover the extent to which *Ss* may have provided labels of their own invention during the pre-training procedure. It will probably not be possible to evaluate this type of training method adequately until measures of this factor are included as part of the data in experiments using directed-attention training.

*Performance set.* Recently experimenters in this area have been concerned that the positive transfer ob-

tained in predifferentiation experiments might, at least in part, be due to the transfer of such general factors as "warm-up" or "learning-how-to-learn." The term "performance set" is used here to designate all such factors. In general, concern over factors of this sort has been evidenced by the inclusion of special control groups which provide the possibility of measuring their effect. Most often the training of these groups is of the Irrelevant-S sort, although occasionally simple familiarization training has been used. The results appear to be unequivocal. In every case in which this type of training has been compared with verbal paired-associates training of the Relevant-S type it has yielded significantly less transfer. J. H. Cantor (14) and Smith and Goss (36) found it to be equivalent to no pretraining at all. It seems clear that transfer of predifferentiation cannot be accounted for in terms of general factors of this sort.

Summarizing the evidence for and against the various hypotheses which have been offered to account for the transfer of predifferentiation training leads to the following conclusions: (a) The *acquired distinctiveness of cues* and *reduction of intralist generalization* hypotheses imply, in general, the same kinds of operations and lead to the same kinds of predictions. Specific tests of deductions from these hypotheses have failed, however, to receive any experimental support. (b) In its simplest form the *increased meaningfulness* hypothesis is also operationally equivalent to the first two in that the increased meaningfulness of the stimulus is presumed to result from the acquisition of a label (association). It is a corollary of this hypothesis that an increase in positive transfer should result from increasing the meaningfulness of the label itself. This corollary has not re-

ceived unequivocal experimental support. (c) All three of the foregoing hypotheses may logically be compared with the *attention to cues* hypothesis, which states that the learning of a verbal label is not an essential part of the predifferentiation process. This sort of comparison may properly be made only if some control is effected over the tendency for Ss to provide labels of their own choosing during the pretraining session. (d) On the basis of the experimental evidence available the transfer of predifferentiation training cannot be accounted for on the basis of transfer of general factors such as "warm-up" or "learning-how-to-learn."

#### OTHER CONSIDERATIONS

The primary conclusion to be derived from this survey of experiments on the transfer of predifferentiation training is that the hypotheses which have so far been offered to account for the transfer phenomenon are all unsatisfactory. They all appear to be stated in testable terms; yet, with one exception, there seems to be no experimental basis for choosing among them. When a reproducible effect, such as the production of positive transfer through predifferentiation training, can be accounted for on the basis of a variety of equally plausible hypotheses, it is likely that all of the hypotheses are dealing with superficial aspects of the situation. It becomes necessary, then, to re-examine the whole problem to determine wherein we have failed to discern the crucial factors, to manipulate the most relevant variables, and to organize our thinking along the most appropriate conceptual dimensions. We must determine whether the superficiality of our thinking is due to a failure of observation or of definition.

With these objectives in mind, let

us examine once again the various hypotheses now current. Three of the four remaining hypotheses may be discussed together: *acquired distinctiveness of cues*, *reduction in intralist generalization*, and *attention to cues*. The first question we may ask of these hypotheses is: What is a cue? It is strongly implied that the word *cue* refers to a stimulus characteristic which may be independently varied. Examination of the experiments generated by these hypotheses, however, reveals that the existence of cues is characteristically inferred from the fact that learning produces more reliable discrimination responses. There is no objection to making such an inference, but it can be argued that there are more efficient ways of investigating the importance of cues in learning and transfer. For example, Kurtz (26) recently showed that either positive or negative transfer could be obtained, depending upon the presence or absence in the transfer task stimuli of the particular cues which had provided the basis for discrimination during the pretraining. Kurtz used two-dimensional forms as his stimuli, and it is not too difficult to manipulate cues objectively in stimuli of this sort. What, though, is a cue when the stimulus is verbal? Is the cue a letter, a pattern of letters, a sound, or perhaps the connotations of the verbal symbol? Any or all of these may be cues, and some experimenters have used these definitions explicitly while others have failed to specify which definition was being used. Needless to say, we cannot determine whether any "cue" has acquired distinctiveness unless we know precisely what is meant by a cue.

An examination of the concept of "distinctiveness" (a term which Gibson [19] has also used in connection with reduction in intralist generaliza-

tion) leads to many of the same questions. Acquired distinctiveness can be inferred from positive transfer or from a reduction in intralist errors, but it is potentially a more powerful explanatory concept when measured independently of the phenomenon it is designed to explain. The writer has previously pointed out (1, 3) that distinctiveness is a stimulus attribute which can be measured by psychophysical methods, and that an increase in distinctiveness should be accompanied by a change in the threshold for discrimination. It has been shown that an increase in distinctiveness (in this sense) follows upon verbal paired-associates training when the perceptual test is one of delayed recognition (3) but not when the test is one of same-different discrimination (1). These results imply that the acquired distinctiveness affects not the perception of the stimulus but the memory of it, which is consistent with the results obtained by Lawrence and Coles (27).

As before, it is easier to discuss these concepts in connection with form stimuli than in connection with verbal stimuli. It is hard to imagine the ways in which a verbal stimulus becomes more distinctive until it is decided what the cues for recognition are. Likewise, the usefulness of an explanation based upon *attention* to cues will be slight until it is possible to specify more adequately what attention is and to make some guesses about how it operates to facilitate recognition or memory.

The hypothesis that transfer of predifferentiation training results from increasing the *meaningfulness* of the stimuli derives essentially from the core-context theory of meaning and has not had the sort of formal theoretical development that exists for the more behavioristic hypotheses of Miller and Dollard (29) and Gib-

son (18). No mechanism for accomplishing the transfer has been postulated beyond the simple assumption that more meaningful stimuli are more easily learned and more easily remembered. Even within this simple conceptualization, however, there remain many unanswered questions. What, precisely, is meant by the *meaningfulness* of a stimulus? Attempts have been made to quantify the meaningfulness of verbal materials (Glaze, [20], Noble [31]), and attempts are currently in progress to develop meaningfulness scales for nonsense forms, but it remains questionable whether these measures so far developed will be adequate to account for predifferentiation transfer. What are the differential effects on learning of stimulus meaning and response meaning, when meaning is defined as the number, intensity, or latency of associations? And, parenthetically, are these three definitions of meaning equivalent? What would be the effect on transfer of requiring *S* to learn several responses to each stimulus, each to a partial criterion; would this be as effective as learning one response thoroughly? What is the relation between *meaningfulness*, defined by the associations elicited by a single term, such as a stimulus or a response, and *belongingness*, defined as an existing connotative or denotative relationship between a pair of terms either stimuli, responses, or both? Are these equivalent with respect to their effect on transfer? Most of these questions are susceptible to experimental test on the basis of concepts presently available, but it is the writer's belief that no real understanding of them will be achieved until the whole problem of meaning has been more satisfactorily resolved.

The foregoing discussion does not nearly exhaust the questions which

could be asked concerning the hypotheses which have so far been proposed, but they should indicate that all these hypotheses have described the stimulus-response situation in terms which are superficially plausible but which are not easily quantified or, in some cases, even operationally specifiable. To state the case in the most obvious terms, stimuli are not usually simple events which can be described as "lights," "forms," or "words," nor can responses be described in terms equally simple. Stimuli and responses are not independent entities which can be adequately described solely in terms of themselves, but rather they are always members of a class whose size and class-characteristics are a function of the total experience of the individual subject. Naturally, psychology cannot hope to deal with the total apperceptive mass of each subject in relation to each stimulus and each response, and usually it is not necessary to do so. It is possible, however, to deal with smaller classes which are highly relevant. For example, in an experiment on paired-associates learning the stimuli and responses to be learned constitute important classes, and the individual items derive important attributes from the fact that they are members of these classes.

A recent experiment by Attneave (6) provides an excellent demonstration of this principle. He was interested in investigating the "schema" hypothesis, which has been proposed by Bartlett (9), Oldfield (33), Woodworth (38), Hebb (24), and others. Attneave required a group of *Ss* to draw repeatedly from memory a prototype nonsense form; he then required them to learn differential responses to a set of nonsense forms which were random variations on the prototype. The group which had

practiced drawing the prototype, which was a "mean" of the variations, was significantly better at the paired-associates learning task than was a group which had practiced on an irrelevant form. The results were interpreted as showing that the memory of the prototype had served as a "schema" about which the variations might be organized and learned.

The results obtained in this experiment are similar to those obtained in usual predifferentiation experiments, but the kind of pretraining used was wholly different. These results cannot be accounted for by any of the hypotheses so far discussed because no formal *predifferentiation* of the stimuli was involved. They pose a problem not only for predifferentiation hypotheses but also for all current conceptualizations about transfer of training.

Attneave suggests that schema learning is always involved in predifferentiation training. In the course of the pretraining the subject learns at least three things about the class of stimuli within which differentiations are being made: (a) the central tendency of the class; (b) how its members may differ from one another; and (c) the dispersion of the class—i.e., how much its members may differ from one another on the several dimensions of variability. While many other things about the stimuli are undoubtedly also learned, these three class parameters together form the "schema" to which the individual stimuli are related.

Looking at the problem of predifferentiation training (and the whole problem of transfer of training) from this point of view leads to an experimental program somewhat different from that which has existed up to now. The primary requirement for such a view is that a thorough knowledge of the stimulus be available.

The discriminable attributes of the stimulus must be quantitatively related both to its physical structure and to various experiential factors. Research of this sort is already in progress on both verbal (31) and non-verbal (2, 5) materials. When it is possible to describe stimuli in these terms it will be possible to manipulate the conditions of an experiment in such a way that the specific factors responsible for transfer can be identified.

It is not necessarily true that the kinds of hypotheses generated by experiments of this sort will be very different from the kinds currently available. The difference will be that the sorts of hypothetical constructs and intervening variables which will be

formulated will be based on detailed knowledge of the functional relationships between the discriminable stimulus attributes, on the one hand, and structural and experiential factors on the other. The fact that it is somewhat easier to obtain such functional relationships in the case of nonverbal than of verbal stimuli (4) suggests that nonsense forms may come to be preferred over nonsense syllables as the ideal stimuli for transfer studies. In any case, it is the thesis of this discussion that more satisfactory hypotheses to account for transfer effects will be developed only when it becomes possible to give a more adequate quantitative description of the stimulus.

## REFERENCES

- ARNOULT, M. D. Transfer of predifferentiation training in simple and multiple shape discrimination. *J. exp. Psychol.*, 1953, **45**, 401-409.
- ARNOULT, M. D. Familiarity and recognition of nonsense shapes. *J. exp. Psychol.*, 1956, **51**, 269-276.
- ARNOULT, M. D. Recognition of shapes following paired-associates pretraining. In G. Finch & F. Cameron (Eds.), *Symposium on Air Force human engineering, personnel, and training research*. Washington, D. C.: National Academy of Sciences—National Research Council, Publication 455, 1956. Pp. 1-9.
- ATTNEAVE, F., & ARNOULT, M. D. Methodological considerations in the quantitative study of shape and pattern perception. *Psychol. Bull.*, 1956, **53**, 452-471.
- ATTNEAVE, F. Physical determinants of the judged complexity of shapes. *J. exp. Psychol.*, 1957, **53**, 221-227.
- ATTNEAVE, F. Transfer of experience with a class-schema to identification-learning of patterns and shapes. *J. exp. Psychol.*, 1957, **54**, in press.
- BAKER, K. E., & WYLIE, R. C. Transfer of verbal training to a motor task. *J. exp. Psychol.*, 1950, **40**, 632-638.
- BALDWIN, R. D. Discrimination learning as a function of stimulus predifferentiation and mediated association training. Unpublished doctor's dissertation, State Univer. of Iowa, 1954.
- BARTLETT, F. C. *Remembering, a study in experimental and social psychology*. Cambridge, Eng.: Cambridge Univer. Press, 1932.
- BATTIG, W. F. Transfer from verbal pre-training to motor performance as a function of motor task complexity. *J. exp. Psychol.*, 1956, **51**, 371-378.
- BIRGE, J. S. The role of verbal response in transfer. Unpublished doctor's dissertation, Yale Univer., 1941.
- CAMPBELL, V., & FREEMAN, J. T. Some functions of experimentally-induced language in perceptual learning. *Percept. Mot. Skills*, 1955, **1**, 71-79.
- CANTOR, G. N. Effects of three types of pretraining on discrimination learning in preschool children. *J. exp. Psychol.*, 1955, **49**, 339-342.
- CANTOR, J. H. Amount of pretraining as a factor in stimulus predifferentiation and performance set. *J. exp. Psychol.*, 1955, **50**, 180-184.
- DYSINGER, D. W. An investigation of stimulus predifferentiation in a choice discrimination problem. Unpublished doctor's dissertation, State Univer. of Iowa, 1951.
- GAGNÉ, R. M., & BAKER, KATHERINE E. Stimulus predifferentiation as a factor in transfer of training. *J. exp. Psychol.*, 1950, **40**, 439-451.
- GERJUOY, I. R. Discrimination learning as a function of the similarity of the

could be asked concerning the hypotheses which have so far been proposed, but they should indicate that all these hypotheses have described the stimulus-response situation in terms which are superficially plausible but which are not easily quantified or, in some cases, even operationally specifiable. To state the case in the most obvious terms, stimuli are not usually simple events which can be described as "lights," "forms," or "words," nor can responses be described in terms equally simple. Stimuli and responses are not independent entities which can be adequately described solely in terms of themselves, but rather they are always members of a class whose size and class-characteristics are a function of the total experience of the individual subject. Naturally, psychology cannot hope to deal with the total apperceptive mass of each subject in relation to each stimulus and each response, and usually it is not necessary to do so. It is possible, however, to deal with smaller classes which are highly relevant. For example, in an experiment on paired-associates learning the stimuli and responses to be learned constitute important classes, and the individual items derive important attributes from the fact that they are members of these classes.

A recent experiment by Attneave (6) provides an excellent demonstration of this principle. He was interested in investigating the "schema" hypothesis, which has been proposed by Bartlett (9), Oldfield (33), Woodworth (38), Hebb (24), and others. Attneave required a group of Ss to draw repeatedly from memory a prototype nonsense form; he then required them to learn differential responses to a set of nonsense forms which were random variations on the prototype. The group which had

practiced drawing the prototype, which was a "mean" of the variations, was significantly better at the paired-associates learning task than was a group which had practiced on an irrelevant form. The results were interpreted as showing that the memory of the prototype had served as a "schema" about which the variations might be organized and learned.

The results obtained in this experiment are similar to those obtained in usual predifferentiation experiments, but the kind of pretraining used was wholly different. These results cannot be accounted for by any of the hypotheses so far discussed because no formal *predifferentiation* of the stimuli was involved. They pose a problem not only for predifferentiation hypotheses but also for all current conceptualizations about transfer of training.

Attneave suggests that schema learning is always involved in predifferentiation training. In the course of the pretraining the subject learns at least three things about the class of stimuli within which differentiations are being made: (a) the central tendency of the class; (b) how its members may differ from one another; and (c) the dispersion of the class—i.e., how much its members may differ from one another on the several dimensions of variability. While many other things about the stimuli are undoubtedly also learned, these three class parameters together form the "schema" to which the individual stimuli are related.

Looking at the problem of predifferentiation training (and the whole problem of transfer of training) from this point of view leads to an experimental program somewhat different from that which has existed up to now. The primary requirement for such a view is that a thorough knowledge of the stimulus be available.

The discriminable attributes of the stimulus must be quantitatively related both to its physical structure and to various experiential factors. Research of this sort is already in progress on both verbal (31) and non-verbal (2, 5) materials. When it is possible to describe stimuli in these terms it will be possible to manipulate the conditions of an experiment in such a way that the specific factors responsible for transfer can be identified.

It is not necessarily true that the kinds of hypotheses generated by experiments of this sort will be very different from the kinds currently available. The difference will be that the sorts of hypothetical constructs and intervening variables which will be

formulated will be based on detailed knowledge of the functional relationships between the discriminable stimulus attributes, on the one hand, and structural and experiential factors on the other. The fact that it is somewhat easier to obtain such functional relationships in the case of nonverbal than of verbal stimuli (4) suggests that nonsense forms may come to be preferred over nonsense syllables as the ideal stimuli for transfer studies. In any case, it is the thesis of this discussion that more satisfactory hypotheses to account for transfer effects will be developed only when it becomes possible to give a more adequate quantitative description of the stimulus.

## REFERENCES

- ARNOULT, M. D. Transfer of predifferentiation training in simple and multiple shape discrimination. *J. exp. Psychol.*, 1953, **45**, 401-409.
- ARNOULT, M. D. Familiarity and recognition of nonsense shapes. *J. exp. Psychol.*, 1956, **51**, 269-276.
- ARNOULT, M. D. Recognition of shapes following paired-associates pretraining. In G. Finch & F. Cameron (Eds.), *Symposium on Air Force human engineering, personnel, and training research*. Washington, D. C.: National Academy of Sciences—National Research Council, Publication 455, 1956. Pp. 1-9.
- ATTNEAVE, F., & ARNOULT, M. D. Methodological considerations in the quantitative study of shape and pattern perception. *Psychol. Bull.*, 1956, **53**, 452-471.
- ATTNEAVE, F. Physical determinants of the judged complexity of shapes. *J. exp. Psychol.*, 1957, **53**, 221-227.
- ATTNEAVE, F. Transfer of experience with a class-schema to identification-learning of patterns and shapes. *J. exp. Psychol.*, 1957, **54**, in press.
- BAKER, K. E., & WYLIE, R. C. Transfer of verbal training to a motor task. *J. exp. Psychol.*, 1950, **40**, 632-638.
- BALDWIN, R. D. Discrimination learning as a function of stimulus predifferentiation and mediated association training. Unpublished doctor's dissertation, State Univer. of Iowa, 1954.
- BARTLETT, F. C. *Remembering, a study in experimental and social psychology*. Cambridge, Eng.: Cambridge Univer. Press, 1932.
- BATTIG, W. F. Transfer from verbal pre-training to motor performance as a function of motor task complexity. *J. exp. Psychol.*, 1956, **51**, 371-378.
- BIRGE, J. S. The role of verbal response in transfer. Unpublished doctor's dissertation, Yale Univer., 1941.
- CAMPBELL, V., & FREEMAN, J. T. Some functions of experimentally-induced language in perceptual learning. *Percept. Mot. Skills*, 1955, **1**, 71-79.
- CANTOR, G. N. Effects of three types of pretraining on discrimination learning in preschool children. *J. exp. Psychol.*, 1955, **49**, 339-342.
- CANTOR, J. H. Amount of pretraining as a factor in stimulus predifferentiation and performance set. *J. exp. Psychol.*, 1955, **50**, 180-184.
- DYSINGER, D. W. An investigation of stimulus predifferentiation in a choice discrimination problem. Unpublished doctor's dissertation, State Univer. of Iowa, 1951.
- GAGNÉ, R. M., & BAKER, KATHERINE E. Stimulus predifferentiation as a factor in transfer of training. *J. exp. Psychol.*, 1950, **40**, 439-451.
- GERJOUY, I. R. Discrimination learning as a function of the similarity of the

stimulus names. Unpublished doctor's dissertation, State Univer. of Iowa, 1953.

18. GIBSON, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196-229.
19. GIBSON, E. J. Intra-list generalization as a factor in verbal learning. *J. exp. Psychol.*, 1942, **30**, 185-200.
20. GLAZE, J. A. The association value of non-sense syllables. *J. genet. Psychol.*, 1928, **35**, 255-269.
21. GOSS, A. E. Transfer as a function of type and amount of preliminary experience with the task stimuli. *J. exp. Psychol.*, 1953, **46**, 419-428.
22. HAKE, H. W., & ERIKSEN, C. W. Effect of number of permissible response categories on learning of a constant number of visual stimuli. *J. exp. Psychol.*, 1955, **50**, 161-167.
23. HAKE, H. W., & ERIKSEN, C. E. Role of response variables in recognition and identification of complex visual forms. *J. exp. Psychol.*, 1956, **52**, 235-243.
24. HEBB, D. O. Organization of behavior. New York: Wiley, 1949.
25. KURTZ, K. H., & HOVLAND, C. I. The effect of verbalizations during observation of stimulus objects upon accuracy of recognition and recall. *J. exp. Psychol.*, 1953, **45**, 157-164.
26. KURTZ, K. H. Discrimination of complex stimuli: the relationship of training and test stimuli in transfer of discrimination. *J. exp. Psychol.*, 1955, **50**, 283-292.
27. LAWRENCE, D. H., & COLES, G. R. Accuracy of recognition with alternatives before and after the stimulus. *J. exp. Psychol.*, 1954, **47**, 208-214.
28. MCALISTER, D. E. The effects of various kinds of relevant verbal pretraining on subsequent motor performance. *J. exp. Psychol.*, 1953, **46**, 329-336.
29. MILLER, N. E., & DOLLARD, J. *Social learning and imitation*. New Haven: Yale Univer. Press, 1941.
30. MILLER, N. E. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155-178.
31. NOBLE, C. E. An analysis of meaning. *Psychol. Rev.*, 1952, **59**, 421-430.
32. NOBLE, C. E. The meaning-familiarity relationship. *Psychol. Rev.*, 1953, **60**, 89-98.
33. OLDFIELD, R. C. Memory mechanisms and the theory of schemata. *Brit. J. Psychol.*, 1954, **45**, 14-23.
34. ROBINSON, J. S. The effect of learning verbal labels for stimuli on their later discrimination. *J. exp. Psychol.*, 1955, **49**, 112-115.
35. ROSSMAN, I. L., & GOSS, A. E. The acquired distinctiveness of cues: the role of discriminative verbal responses in facilitating the acquisition of discriminative motor responses. *J. exp. Psychol.*, 1951, **42**, 173-182.
36. SMITH, S. L., & GOSS, A. E. The role of the acquired distinctiveness of cues in the acquisition of motor skill in children. *J. genet. Psychol.*, 1955, **87**, 11-24.
37. THORNDIKE, E. L. The fundamentals of learning. New York: Teachers Coll., Columbia Univer., 1932.
38. WOODWORTH, R. S. *Experimental psychology*. New York: Holt, 1938.

*Received for early publication April 9, 1957*

## THREE CRITERIA FOR THE USE OF ONE-TAILED TESTS

HERBERT D. KIMMEL

*University of Southern California*

Examination of the recent literature on the question of when to use one-tailed tests of significance in psychological research reveals a state of unresolved disagreement. A variety of differing opinions (1, 2, 5, 7, 8, 9, 10, pp. 62-63) have been presented, ranging from Burke's (2) exhortation that psychologists should never report one-tailed tests in the public literature to Jones' (8) statement that we may not only do so, but, in certain instances, we will be in error if we fail to do so.

It is by no means necessary for psychologists to agree on all matters of importance to them. Disagreement regarding methodological considerations, however, especially when they bear on how and when propositions shall be accepted as true or rejected as false, should not be permitted to persist indefinitely. The argument is not settled by noting, as Burke (2) does, that the increased use of one-tailed tests may result in the one-tailers scoring a sociological victory almost before the controversy has begun. Actually, this observation by Burke does not coincide completely with the fact that many responsible investigators have continued to employ two-tailed tests (in situations calling for one-tailed tests according to Jones' view) long after the opening of the one-tailed avenue.<sup>1</sup>

<sup>1</sup> An example of an experiment with an explicit directional hypothesis, but employing a two-tailed test, is reported by Davitz (3). This experimenter reasoned that the injection of tetraethylammonium prior to extinction trials would inhibit the punishing effect of the emotional response under study and, consequently, would result in faster extinction in the experimental animals than in a placebo-injected control group. Instead, Davitz found

In attempting to arrive at a set of acceptable criteria for the use of one-tailed tests, it is important to note that the argument is not one of mathematical statistics but primarily one of experimental logic. Burke and Jones would agree that one-tailed tests should be used to test one-tailed hypotheses; their disagreement concerns when one-tailed hypotheses should and should not be made.

Before proceeding to the proposed criteria, it would be of value to consider the difference between one- and two-tailed hypotheses from a viewpoint that has not been stressed by previous writers. All concerned agree that a given mean difference in the hypothesized direction is "more significant"<sup>2</sup> under a one-tailed hypothesis (in the correct direction) than under a two-tailed hypothesis. This is due to the fact that there are exactly twice as many chances of

that the experimental group extinguished slower than the control group, the difference in mean number of trials being significant at the 5 per cent level using a two-tailed test. A one-tailed hypothesis in this experiment (as would have been urged by Jones) would have made it impossible to evaluate the significance of the obtained difference. A study by Hilgard et al. (6), on the other hand, stated a one-tailed hypothesis in a situation in which a difference in the unpredicted direction could have been predicted with as much justification on the basis of previous work. They obtained a difference in their predicted direction that was significant at the 5 per cent level using a one-tailed test. Their rejection of the null hypothesis on the basis of the difference they obtained is the equivalent of loosening the conventional 5 and 1 per cent standards.

<sup>2</sup> That is to say, by chance, the unidirectional event is half as probable as the bidirectional; thus its occurrence, being half as likely, is twice as significant.

committing a type 1 error, with a given mean difference, under a two-tailed hypothesis. The important consideration is that this gain does not accrue without concomitant loss. Even psychology has its law of conservation of energy.

The price that is paid in return for the increased power of one-tailed tests over two-tailed tests stems from the fact that two-tailed null hypotheses are actually more specific than their one-tailed counterparts. A two-tailed null hypothesis can be rejected by a large observed difference in either direction but a one-tailed null hypothesis cannot be rejected by a difference in the unpredicted direction, no matter how large this difference may be. This means that an experimenter using a one-tailed hypothesis *cannot* conclude that an extreme difference in the unpredicted direction is reliably different from zero difference. This limitation cannot be shrugged off by the comment, "We have no interest in a difference in the opposite direction." Scientists are interested in empirical fact regardless of its relationship to their preconceptions.

The meaning of this limitation is exemplified even in applied studies; e.g., those intended to answer the question whether a new product is "better" than the current product. It would be desirable to be able to conclude that the new product is not only "not better" (which is all that failure to reject a one-tailed null hypothesis permits<sup>1</sup>), but, in fact, "poorer." The decision not to market the proposed new product would fol-

low from either conclusion, it is true, but the additional information available as a result of rejecting a two-tailed null hypothesis from the unexpected side could very well indicate a course of behavior quite different from that indicated by the mere inability to reject a specific one-tailed null hypothesis.

It is hoped that the following criteria will be acceptable to psychological investigators as a group and will be adopted conventionally as a guide. The ultimate consequence of our present state of ambiguity on this matter can only be confusion and subsequent retrogression to a more primitive level of scientific communication and understanding.

#### CRITERIA FOR THE USE OF ONE-TAILED TESTS

1. Use a one-tailed test when a difference in the unpredicted direction, while possible, would be psychologically meaningless. An example of this situation might be found in the comparison of experimental and control groups on a skilled task for which only the experimental group has received appropriate training. The experiment would have to be designed in such a way as to eliminate all known conditions that could produce opposite results (e.g., not testing immediately after training to avoid fatigue effects, not testing too long after training to avoid memory loss effects, etc.). Since a difference in the unpredicted direction will have been declared beforehand to have no possible meaning (in terms of previous data and present operations) one-tailed hypotheses could not undergo metamorphosis into two-tailed hypotheses to permit testing the significance of differences in the unpredicted direction.

2. Use a one-tailed test when re-

<sup>1</sup> As Fisher (4) has pointed out, an experimenter never "accepts" the null hypothesis, he merely fails to reject it on the basis of his data. This is one reason why the null hypothesis in a particular experiment should be stated as specifically as possible.

sults in the unpredicted direction will, under no conditions, be used to determine a course of behavior different in any way from that determined by no difference at all. This situation is exemplified by the applied study discussed above, in which a new product is compared with one already on the market.

3. Use a one-tailed test when a directional hypothesis is deducible from psychological theory but results in the opposite direction are not deducible from coexisting psychological theory. If results in the opposite direction are explainable in terms of the constructs of existing theory, no matter how divergent from the experimenter's theoretical orientation this theory may be, the statistical hypothesis must be stated in a way that permits evaluation of opposite results. If this criterion were not already implicitly accepted by psychologists, crucial experiments could never be performed.

It should be apparent that the three criteria stated above are actually slightly differing reflections of the same underlying precept. Neither the ethical nor the logical decisions of individual scientists can be prescribed beforehand by any set of standards, no matter how all-pervasive these standards may seem at a given moment. The three criteria proposed above, however, are offered as temporary guideposts until such time as a new set of temporary criteria supersede them. Proponents of one-tailed tests, such as Jones (7, 8), cannot complain that the use of these criteria will reduce the number of one-tailed tests to near zero, without admitting that these tests have been misused in the past. Opponents of one-tailed tests, such as Burke (1, 2), should welcome this attempt to limit the use of one-tailed tests to those infrequent situations provided for by the proposed criteria.

#### REFERENCES

1. BURKE, C. J. A brief note on one-tailed tests. *Psychol. Bull.*, 1953, **50**, 384-387.
2. BURKE, C. J. Further remarks on one-tailed tests. *Psychol. Bull.*, 1954, **51**, 587-590.
3. DAVITZ, J. R. Decreased autonomic functioning and extinction of a conditioned emotional response. *J. comp. physiol. Psychol.*, 1953, **46**, 311-313.
4. FISHER, R. A. *The design of experiments*. London: Oliver and Boyd, 1947.
5. HICK, W. E. A note on one-tailed and two-tailed tests. *Psychol. Rev.*, 1952, **59**, 316-318.
6. HILGARD, E. R., JONES, L. V., & KAPLAN, S. J. Conditioned discrimination as related to anxiety. *J. exp. Psychol.*, 1951, **42**, 94-99.
7. JONES, L. V. Tests of hypotheses: One-sided vs. two-sided alternatives. *Psychol. Bull.*, 1952, **49**, 43-46.
8. JONES, L. V. A rejoinder on one-tailed tests. *Psychol. Bull.*, 1954, **51**, 585-586.
9. MARKS, M. R. Two kinds of experiment distinguished in terms of statistical operations. *Psychol. Rev.*, 1951, **58**, 179-184.
10. McNEMAR, Q. *Psychological statistics*. New York: Wiley, 1955.

Received February 21, 1957.

## SLEEP, WAKEFULNESS, AND CONSCIOUSNESS

NATHANIEL KLEITMAN

*Department of Physiology, University of Chicago*

In Ellingson's extensive review of "Brain Waves and Problems of Psychology" (7) there is a small section devoted to "Sleep and Wakefulness," with a well-considered and meticulously worded summary of recent work on the interrelations between the brain-stem reticular formation (BSRF) and the cerebral cortex, as revealed by EEG studies. This summary was criticized by Schmidt (13), with particular reference to the following two statements: "It is clear from these results that the BSRF is essential to the maintenance of the waking state under normal conditions," and "Taken together these findings indicate that a background of maintained activity in the BSRF accounts for the maintenance of wakefulness, while reduction of its activity precipitates a state of somnolence or unconsciousness." Schmidt contended that "various observations do not support the notion that reduction of activity of the reticular formation must result in behavioral sleep."

What were these "various observations?" They pertained to the effect of atropine on the EEG pattern of cats, rabbits, and dogs. Among others, Bradley and Elkes (4), working on unrestrained unanesthetized cats, found that "Large doses (2 to 3 mg/kg i.p.) of atropine sulphate produced high amplitude waves, interspersed by bursts of fast activity. In general appearance these changes resembled the patterns characteristic of sleep. They did, however, differ from the latter in their failure to show a cortical 'alerting response' to sensory stimuli, although the animal could, in fact, be roused." Rinaldi

and Himwich (11) noted that in unanesthetized, but curarized, rabbits, after "high doses of atropine, the electrocorticogram shows a stable pattern of sleep that cannot be modified by stimulation of any sort. It is impossible to produce an alert pattern and thus desynchronize the sleeping potentials." Most significantly, Wikler's dogs (15), after an injection of atropine, "were definitely 'excited' and had to be restrained to permit recording, at a time when the 'sleep patterns' were evident in the EEG tracings. When released, these atropinized animals jumped off the table and spontaneously returned to the animal quarters in the laboratory." From these atropine effects Schmidt concluded that "behavioral wakefulness can accompany reduction of activity in the BSRF." This conclusion is unjustified, as the studies referred to by Schmidt pertained only to the influence of the BSRF and the effects of peripheral stimulation on the cerebral cortex, and not at all on the reduction of activity in the BSRF with respect to lower centers. Schmidt summarizes his criticism in two sentences: "The pharmacological data cited here indicate inactivity of the BSRF by itself is an insufficient condition for the occurrence of sleep" and "Consequently, sleep and the so-called 'sleep pattern' are not necessarily correlated." It would be more appropriate for Schmidt to interchange his premise and conclusion to read about as follows: "Behavioral sleep and the so-called 'sleep pattern' are not necessarily correlated. Consequently, the pharmacological data cited here have no bearing on the question

whether the activity of the BSRF is essential to the maintenance of the waking state."

What prompts me to reopen this discussion is not so much Schmidt's criticisms, which appear to be a result of confusion over the operation of the BSRF, as Ellingson's retraction (8) of his statement that "the BSRF is essential to the maintenance of the waking state." Taking into consideration that the critic's name was not Lysenko and that he was not backed by the might of a totalitarian party, Ellingson's response shows that the state of semantic confusion was not limited to Schmidt. Furthermore, the second sentence to which he took exception introduces an element unchallenged by Schmidt, but disturbing to me. Ellingson spoke of the reduction of activity of the BSRF as precipitating "a state of somnolence or unconsciousness." He further stated, in his original review (7), that "it is a moot question whether consciousness as a psychological state, in man at least, is possible without the cortex." And in his reply to Schmidt, Ellingson (8) added that he "did not intend to give the impression that the reticular formation is by itself responsible for the state of consciousness." Thus, the terms consciousness and unconsciousness are introduced into a discourse on wakefulness and sleep, compounding the confusion. Also, Schmidt, in the very beginning of his note (13), probably inadvertently, imputed to Ellingson a conclusion that "the brain stem reticular formation is identical with Kleitman's 'waking center.'" I was careful to point out (9) that the subcortical system in question should be designated as the "wakefulness center," and not the "waking center," as it is responsible for the maintenance of wakefulness, and not merely for the phenomenon of arousal from sleep,

and Ellingson himself had at no time used the term "waking center." However, in his reply to Schmidt, Ellingson (8) defended his use of the term "wakefulness center" and indicated that, in his review (7), he came to look upon the BSRF as "a wakefulness center," rather than "the wakefulness center," adding that "perhaps it was unwise to retain it [the term] thereafter, even while substituting the article 'a' for 'the.'" If there is more than one wakefulness center, where would Ellingson place them? And what is the relation of consciousness to wakefulness and sleep?

#### SLEEP AND WAKEFULNESS

Apart from the Schmidt-Ellingson controversy, and in more general terms, one may ask: how have the EEG and other findings affected the acceptability of the evolutionary theory of sleep and wakefulness which I propounded nearly two decades ago? At the time, I disclaimed exclusive authorship for this theory, indicating that I intended "to draw freely on the experimental results and theoretical considerations of others, keeping in mind the requirement that a theory must be in agreement with known facts and should, if possible, be susceptible of experimental verification" (9). I further stated that I had "no doubt that modifications will be required in this theory as new facts are brought to light." By the scheme proposed, an innate state of wakefulness, designated as "wakefulness of necessity," is maintained through the activity of a subcortical wakefulness "center," which, because of the current aversion to the conception of centers, will hereafter be called the mesodiencephalic wakefulness system (MDWS). This system operates in the absence or incapacitation of the cerebral cortex, employing feed-

back circuits with lower regions of the nervous system and peripheral receptors and effectors. Fatigue or a cyclical decrease in activity of the MDWS leads to sleep. There is nothing to learn or forget individually: phylogenetic development and ontogenetic maturation account for the alternation of the innate sleep and wakefulness. It can be seen in newborn infants and in older anencephalous children, as well as in decorticated higher mammals. There are two criteria for the passage from innate wakefulness to innate sleep: (a) a marked decrease in activity of the skeletal musculature, with the assumption of a characteristic posture, and (b) a raised threshold of reflex excitability. The temporal aspects of the innate sleep-wakefulness periodicity are: (a) a cycle duration (in the human) of 2 to 4 hours, bearing little or no relation to the alternation of night and day, and (b) a dominance of the sleep phase, with, in newborn infants, a sleep-wakefulness ratio of 2:1. The cycle is adjusted to the organism's need for food and water and is essentially a hunger-and-thirst periodicity. It may even be said that the MDWS embraces or is coextensive with centers or systems for fulfilling the general body needs or animalistic functions. Recent observations in our laboratory and examination of data obtained by other investigators (3) revealed the existence, in infants, of a still shorter primitive rest-activity cycle, discernible even during continuous sleep, with a periodicity of 50 to 60 minutes. On a self-demand infant feeding schedule, the interfeeding periods are usually an integer of these primitive cycles, suggesting that if the infant is not awakened, through internal or external stimuli, in the shallow phase of one cycle, it is not likely to awaken till this shallow phase recurs. Variations in illumination, noise level, and ambient temperature are responsible for the early manifestations of nyctohemeral (diurnal) differences, the interfeeding periods, largely spent in sleep, ranging from 1 to 3 cycles in the daytime and from 3 to 5 cycles at night. The mechanism of the primitive cycle is unknown, but, like the cardiac and respiratory cycles, it tends to lengthen with age. It may be a metabolic variation, a pacemaker discharge, or a fatigue-recovery phenomenon.

Grafted on the innate sleep-wakefulness periodicity, as a result of ontogenetic development of the cerebral cortex and of individual experience of the infant, is a new sleep-wakefulness rhythm (9, 10), whose characteristics are: (a) a consolidation of the sleep and wakefulness phases, with a fixed adjustment to the astronomical alternations of night and day and a social acculturation to the family and community pattern of living, long unbroken sleep occurring at night, and (b) a lengthening of the wakefulness phase, which gradually achieves temporal dominance, with a sleep-wakefulness ratio, in man, reversed, becoming 1:2. The addition of the acquired "wakefulness of choice" to the innate "wakefulness of necessity" carries with it a fourfold increase in wakefulness-capacity, the adult human "paying" for each hour of wakefulness with one-half hour of sleep, instead of with two hours, as does the neonate. The acquired, consolidated, once-in-24-hours, night sleep differs from the innate type in the occurrence of dreaming; it resembles innate sleep in the persistence of the primitive rest-activity cycles, which in the adult are of 80 to 90 minutes' duration, manifesting themselves in oscillations in the depth of sleep (6). Whether the primitive periodicity also expresses itself in fluc-

tuations of alertness during the long hours of acquired wakefulness has not yet been established, but the postprandial nap habit of some persons and the satisfaction that others get from a 15-minute cat nap during the late afternoon or early evening letdown suggest a retention of the primitive periodicity in wakefulness.

The mechanism of the acquired sleep-wakefulness rhythm can only be guessed at. It is probably partly nervous, a type of conditioning, and partly endocrine, a hypophyseal-adrenocortical tide, as seen in the nyctohemeral variation in the eosinophile count. The "nervous" component involves the establishment of feed-back circuits between the MDWS and the cerebral cortex, as an extension of those which originally existed between the MDWS and lower centers. Thus, the MDWS can now influence, and be influenced by, structures lying "above" as well as "below." As long as its connections with the cortex are unbroken, the discharges from the active MDWS maintain a "wakefulness" EEG pattern. But if these connections are anatomically severed or pharmacologically or electrically blocked, it should be possible to record a sleep EEG pattern, in the face of overt behavioral wakefulness. Atropine-poisoned dogs may be awake, in spite of their sleep EEG, and decorticated dogs and cats show periodic alternations of sleep and wakefulness in the complete absence of a cerebral cortex. It is curious that in the voluminous literature on the BSRF and the significance of the arousal reaction, no mention is made of the influence of the BSRF downward, as revealed in the behavior of decorticated animals, anencephalous babies and normal human neonates. Desynchronization of cortical waves does occur in awakening, but only if there is a cortex!

It may be mentioned that sleep and wakefulness have recently been observed in completely decorticated monkeys (16), closing the breach between data on man and on subprimate mammals.

Just as a "sleep" EEG is without diagnostic significance in the presence of behavioral wakefulness, as pointed out in the discussion of Schmidt's criticisms of Ellingson's statements, so is a "wakefulness" EEG, obtained during behavioral sleep. Such an EEG was invariably observed in our laboratory (6) during dreaming, with the subjects unquestionably asleep. The dreaming EEG pattern is that of light sleep, a "modified" alpha rhythm, 1 or 2 c.p.s. slower than the subject's wakefulness alpha, and somewhat less regular, but, significantly, without spindling. By the classification of Simon and Emmons (14), such a dreaming EEG pattern would be designated as *A*, a deep drowsy state, or *B*, a transition state between wakefulness and sleep. It appears, then, that the EEG is of value in differentiating degrees of alertness in wakefulness or the depth of sleep, provided wakefulness and sleep are first established by the application of behavioral criteria. In a conflict, behavior should take precedence over the EEG.

#### WAKEFULNESS AND CONSCIOUSNESS

It will be recalled that Schmidt's note (13) dealt only with sleep, but Ellingson, in his reply (8), stated that he "did not intend to give the impression that the reticular formation is by itself responsible for the state of consciousness." This raises the question of the place of consciousness in the sleep-wakefulness dichotomy. The semantic confusion that prevails over the meaning of the term consciousness can be detected in the Transactions of the five annual Macy Foun-

dation Conferences on *Problems of Consciousness* (1), the several conflicting definitions of the term during the symposium on *Brain Mechanisms and Consciousness* (5), and in Schiller's "reconsideration" of the subject (12). Suffice it to say that the term consciousness has often been equated with wakefulness, by myself and others. True, some aspects of consciousness are also found in wakefulness, but only in acquired wakefulness, which depends on the cerebral cortex for its development and maintenance. Sleep and wakefulness, as states, whether innate or acquired, can be objectively observed and to some extent measured—they can be compared and contrasted. Like ice and water, they can be distinguished from each other by simple inspection. The melting point of ice or freezing point of water correspond to the drowsiness level or intermediate stage between wakefulness and sleep. Liquid water may be near the freezing point or close to the boiling point, and so may alertness vary from semi-wakefulness to manic hyperactivity ("boiling mad"). Conversely, the depth of sleep, like the coldness of ice, may be close to the transition state or way down near coma. In consciousness the sleep-wakefulness dichotomy is absent. There is only one state, whose criteria are partly objective, but mainly subjective, the observer making only inferences. These criteria are: (a) critical, as against stereotyped, reactivity, involving an analysis of incoming impulses in the light of one's individual experience and the elaboration of appropriate reactions (thinking), and (b) subsequent spontaneous or evoked recall of events (memory). The level of consciousness is variable and, at any moment, is determined by the degree of one's ability to utilize his past and contribute to his future. In the new-

born infant, or older anencephalous child, as in the decorticated dog or cat, the level of consciousness is close to, if not at, zero. Their responses to stimuli do not meet the criteria for consciousness. Yet they show definite alternation of sleep and wakefulness, of the innate type, to be sure. In animals naturally endowed with a cerebral cortex consciousness manifests itself only in the presence of a functioning cortex, and the same, as already stated, applies to acquired wakefulness. However, consciousness and even acquired wakefulness are not synonymous. Whereas a superior degree of alertness is probably concomitant with a high level of consciousness, and profound slumber may be close to zero of consciousness, certain intermediate levels of consciousness, characterized by rather uncritical reactivity and poor retention for future use, or a short-lasting recall of events, are compatible with either wakefulness or sleep. In delirium, fugues, icteral or posticteral automatism of psychomotor epilepsy, a person may be judged to be behaviorally awake, but his level of consciousness is very low, and he may have a complete amnesia of events. By contrast, an individual behaviorally asleep, but spinning a complicated yarn of a dream, may reach a higher level of consciousness, on the basis of the organization of the dream pattern and a spontaneous recall of events that made up the dream episode. But that is only by contrast, for, as a rule, the level of consciousness in dreaming is lower than in normal wakefulness; the reactivity is less critical and recall poorer. The shorter memory span explains why some persons can honestly maintain that they never, or very seldom, dream. If awakened during dreaming, they not only confirm the fact, but can relate the dream content, but if questioned

upon awakening in the morning, they manifest a complete amnesia.

Aside from his excursions into the semantically treacherous realm of consciousness, Ellingson's statement concerning the dependence of behavioral wakefulness on the activity of the mesodiencephalic wakefulness system (MDWS), topographically coinciding with, or embracing, the brain-stem reticular formation (BSRF), is in accord with presently

available facts, and in conflict with none. His statement is further in accord with the evolutionary theory of sleep and wakefulness. The place of consciousness in this scheme of things is less secure, as there is no commonly accepted definition or even description of consciousness. However, consciousness is absent in innate wakefulness, present in acquired wakefulness, and may be present in acquired sleep, during dreaming.

#### REFERENCES

1. ABRAMSON, H. S. (Ed.). *Problems of consciousness*. New York: Josiah Macy Foundation, 1951-1955.
2. ASERINSKY, E., & KLEITMAN, N. Regularly occurring periods of eye motility, and concomitant phenomena, during sleep. *Science*, 1953, **118**, 273-274.
3. ASERINSKY, E., & KLEITMAN, N. A motility cycle in sleeping infants as manifested by ocular and gross bodily activity. *J. appl. Physiol.*, 1955, **8**, 11-18.
4. BRADLEY, P. B., & ELKES, J. The effect of atropine, physostigmine, and neostigmine on the electrical activity of the brain of the cat. *J. Physiol.*, 1953, **120**, 14P-15P.
5. DELAFRESNAYE, J. F. (Ed.). *Brain mechanisms and consciousness*. Springfield: Thomas, 1954.
6. DEMENT, W., & KLEITMAN, N. Incidence of eye motility during sleep in relation to varying EEG pattern. *Fed. Proc.*, 1955, **14**, 37.
7. ELLINGSON, R. J. Brain waves and problems of psychology. *Psychol. Bull.*, 1956, **53**, 1-34.
8. ELLINGSON, R. J. Comments on Schmidt's "The reticular formation and behav-
- ioral wakefulness." *Psychol. Bull.*, 1957, **54**, 76-78.
9. KLEITMAN, N. *Sleep and wakefulness*. Chicago: Univer. of Chicago Press, 1939.
10. KLEITMAN, N., & ENGELMANN, T. G. Sleep characteristics of infants. *J. appl. Physiol.*, 1953, **6**, 269-282.
11. RINALDI, F., & HIMWICH, H. E. Alerting responses and action of atropine and cholinergic drugs. *A.M.A. Arch. Neurol. Psychiat.*, 1955, **73**, 387-395.
12. SCHILLER, F. Consciousness reconsidered. *A.M.A. Arch. Neurol. Psychiat.*, 1952, **67**, 199-227.
13. SCHMIDT, H., JR. The reticular formation and behavioral wakefulness. *Psychol. Bull.*, 1957, **54**, 75.
14. SIMON, C. W., & EMMONS, W. H. EEG, consciousness, and sleep. *Science*, 1956, **124**, 1066-1069.
15. WIKLER, A. Pharmacologic dissociation of behavior and EEG "sleep pattern" in dogs: morphine, N-allylnormorphine, and atropine. *Proc. Soc. exp. Biol., N. Y.*, 1952, **79**, 261-265.
16. WOOLSEY, C. N. Film presentation and personal communication.

Received March 7, 1957.

## COMMENT ON KLEITMAN'S NOTE

ROBERT J. ELLINGSON  
*Nebraska Psychiatric Institute*

It is obvious that in my comments on Schmidt's note I did not make myself entirely clear. In retracting the statement that "the BSRF is essential to the maintenance of the waking state," all I meant to retract was the word *essential*, on the grounds that it is not proved that structures other than the BSRF are not involved. On the other hand there is no satisfactory evidence that other structures *are* involved. I thought it was implicit in my comments that I did not find Schmidt's evidence wholly convincing; perhaps I should have stated

so explicitly. As a matter of opinion, I feel that there is probably not more than one "wakefulness center," but I do not really know. Until another is identified, I agree that we can very well "make do" with the one we have.

Kleitman's points with regard to the area of semantic confusion surrounding the terms *wakefulness* and *consciousness* are well made. His clarifying discussion is much appreciated.

*Received March 23, 1957.*

## ON WILSON'S DISTRIBUTION-FREE TEST OF ANALYSIS OF VARIANCE HYPOTHESES

QUINN McNEMAR  
*Stanford University*

Since the proposal recently made in this journal by Wilson (7) for a  $\chi^2$  test as a basis for testing hypotheses in 2-way, 3-way, . . . ,  $n$ -way analysis of variance designs has considerable intuitive appeal and entails relatively easy computations, it is apt to be uncritically adopted in lieu of the  $F$  test. The procedure, applicable only to the fixed effects model, involves classifying the scores in each cell as exceeding (or falling below) the over-all median and using the known fact that a total  $\chi^2$ , like a sum of squares, can be apportioned into additive parts. It must be presumed that this proposed test, as all existing distribution-free tests, will be less powerful than the  $F$  test (when the assumptions underlying the latter are met). It is the purpose of this note to contrast the outcome of the Wilson test and the  $F$  test for seven batches of data each involving 2-way classification.

In Table 1 will be found the  $p$  values (juxtaposed) yielded by the  $F$  test and by Wilson's  $\chi^2$  test for row, for column, and for interactive effects. It is difficult to summarize adequately the 21 possible comparisons. If a  $p$  of .01 is used as the level for judging significance, the Wilson test agrees with the  $F$  test for each of the 9 times that  $F$  leads to the acceptance of the null hypothesis. This is exactly what one would expect from a less powerful test, hence such agreement is of little interest. Of the 11  $p$ 's reaching the .01 level by the  $F$  test, 6 fail, 5 by a wide margin, to do so by the Wilson test. Use of the .05 level leads to a similar picture.

Another way of summarizing the comparisons is to note that the median level reached via Wilson's test is .17 in contrast to a median of .01 by the  $F$  test. Also, on the average (median) the  $\chi^2$   $p$ 's are nearly six times larger than the corresponding

TABLE 1  
LEVELS OF SIGNIFICANCE REACHED BY WAY OF  $F$  TEST AND WILSON'S  $\chi^2$  TEST

Source of Data	Number			$p$ Level Reached							
	Rows	Col-	In	Rows		Columns		Interaction			
				umns	Cells	$F$	$\chi^2$	$F$	$\chi^2$		
Edwards (1) p. 209	2	2	10			*	.0002	.005	.53	.05	.06
Lindquist (3) p. 165	3	4	10	.01	.08	.0005**	.005	.20	.43		
McNemar (4) p. 299	2	2	20	.0005	.001	.10	.17	.10	.36		
Snedecor (5) p. 280	4	4	4	.20	.47	.50	.90	.0005	.0004		
Snedecor (5) p. 281	3	2	10	.50	.60	.0005	.01	.10	.10		
Walker and Lev (6) p. 350	4	4	3	.0005	.12	.10	.58	.025	.30		
Artificial data	2	3	20	.005	.03	.0005	.30	.005	.60		

\* The  $F$  for this is 65.19, which is 4 times the  $F$  required for significance at the .0005 level. The corresponding  $t$  of 8.07 would, if interpretable as a normal deviate, be significant at the .000,000,000,000,001 level; but with a  $df$  of 36 it would, perhaps, be significant at only the one in a billion level!

\*\*  $p$  is nearly twice that required for the .0005 level.

*p*'s from the *F* test. A number of startlingly large discrepancies occur.

The foregoing empirical results suggest rather strongly that the power of Wilson's proposed test is very low, whence those who hope to detect effects will not wish to risk its use. It seems unreasonable to believe that any of the six textbook illustrations used violates the assumptions underlying *F*; the last example (artificial data) in Table 1 strictly meets

all the assumptions. For data not satisfying the assumptions of the *F* test, it would be far better to proceed with the *F* test and require that an obtained *F* reach the .01 level in order to be sure of significance at better than the .05 level or that an obtained *F* reach the .005 level for significance at near the .01 or .02 level. These suggestive adjustments are based on the work of Norton, as reported by Lindquist (2).

#### REFERENCES

1. EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.
2. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.
3. LINDQUIST, E. F. *Statistical analysis in educational research*. Boston: Houghton Mifflin, 1940.
4. MCNEMAR, Q. *Psychological statistics*. New York: Wiley, 1955.
5. SNEDECOR, G. W. *Statistical methods*. Ames: Iowa State College Press, 1946.
6. WALKER, H. M., & LEV, J. *Statistical inference*. New York: Holt, 1953.
7. WILSON, K. V. A distribution-free test of analysis of variance hypotheses. *Psychol. Bull.*, 1956, **53**, 96-101.

Received September 24, 1956.

#### ANNOUNCEMENT

Following the publication of the May 1957 issue, ownership of the *Journal of Educational Psychology* will be transferred to the American Psychological Association.

Orders for issues prior to June 1957 should be directed to:

Warwick & York, Inc.  
10 East Centre Street  
Baltimore 2, Maryland

Inquiries on new subscriptions and renewals of subscriptions of issues after July 1957, however, should be directed to:

American Psychological Association  
1333 Sixteenth Street, N. W.  
Washington 6, D. C.

---

**APA Members and Journal Subscribers—Are you going to move?**

***If you move—***

your journals will not follow you from your old address to your new one

***When you move—***

notify the APA Subscription Office

Formerly, journals that could not be delivered because subscribers had not notified the APA of a new address were reclaimed by the APA, and the journal was remailed to the subscriber at his new address. This was always expensive. Recent changes in the postal laws and regulations have made the expense prohibitive. Undeliverable copies are now destroyed by the Post Office. Subscribers who do not receive a journal because of an address change are charged the regular single issue price for a replacement copy.

***So—when you move—***

Notify the postmaster at your old address and guarantee that you will pay the forwarding postage.

Notify the APA Subscription Office as early as possible—by at least the tenth of the month preceding the month when the change should take effect.

---

**AMERICAN PSYCHOLOGICAL ASSOCIATION  
1333 Sixteenth Street N.W.  
Washington 6, D.C.**

---

